

# MÓDULO 1. ESTADÍSTICA DESCRIPTIVA DE UNA VARIABLE



Salvador Naya Fernández. Universidad de A Coruña

# Objetivos del curso

- **Objetivos Generales:** Aprender a utilizar, diseñar y valorar experiencias de enseñanza-aprendizaje con contenidos de estadística y probabilidad.
- **Competencias:** Capacitar al profesor-alumno para diseñar y valorar experiencias de enseñanza-aprendizaje sobre nueva metodología para la enseñanza de la estadística.

# Programa de Probabilidades y estadística en Secundaria (España)

- **Tema 1: Estadística descriptiva unidimensional.** Tipos de variables. Tablas y gráficos. Parámetros estadísticos de centralización, de dispersión y de posición.
- **Tema 2: Distribuciones bidimensionales.** Interpretación de partir de la representación gráfica de una nube de puntos. Grado de relación entre las variables estadísticas. Regresión lineal. Extrapolación de resultados.
- **Tema 3: Azar y probabilidad.** Distribuciones de probabilidad binomial y normal.

# Programa de Probabilidades y estadística en Secundaria (España)

## OBJETIVOS/COMPETENCIAS

- **Utilizar las tablas y gráficas** como instrumento **para el estudio de situaciones empíricas** relacionadas con fenómenos sociales.
- **Interpretar la posible relación** entre las variables de una distribución bidimensional utilizando la recta de regresión y el coeficiente de correlación.
- **Utilizar técnicas estadísticas** elementales para **asignar probabilidades** en situaciones que se ajusten a una distribución de probabilidad binomial o normal.

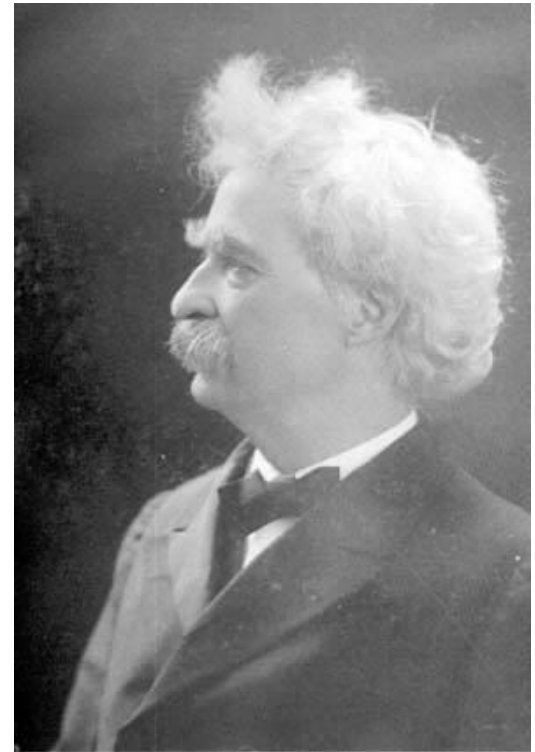
# Programa de Probabilidades y estadística en Secundaria (España)

## CRITERIOS DE EVALUACIÓN

- **Elaborar e interpretar tablas y gráficos estadísticos**, así como calcular parámetros estadísticos más usuales, correspondientes a variables estadísticas discretas y continuas, interpretarlos y valorar cualitativamente la representatividad de las muestras utilizadas.
- ▣ Evaluar la capacidad para interpretar, analizar y valorar informaciones estadísticas presentes en los medios de comunicación.

□ *There are three kinds of lies: lies, dammed lies and statistics.*

*Autobiografía de Mark Twain*



**Samuel Langhorne Clemens**

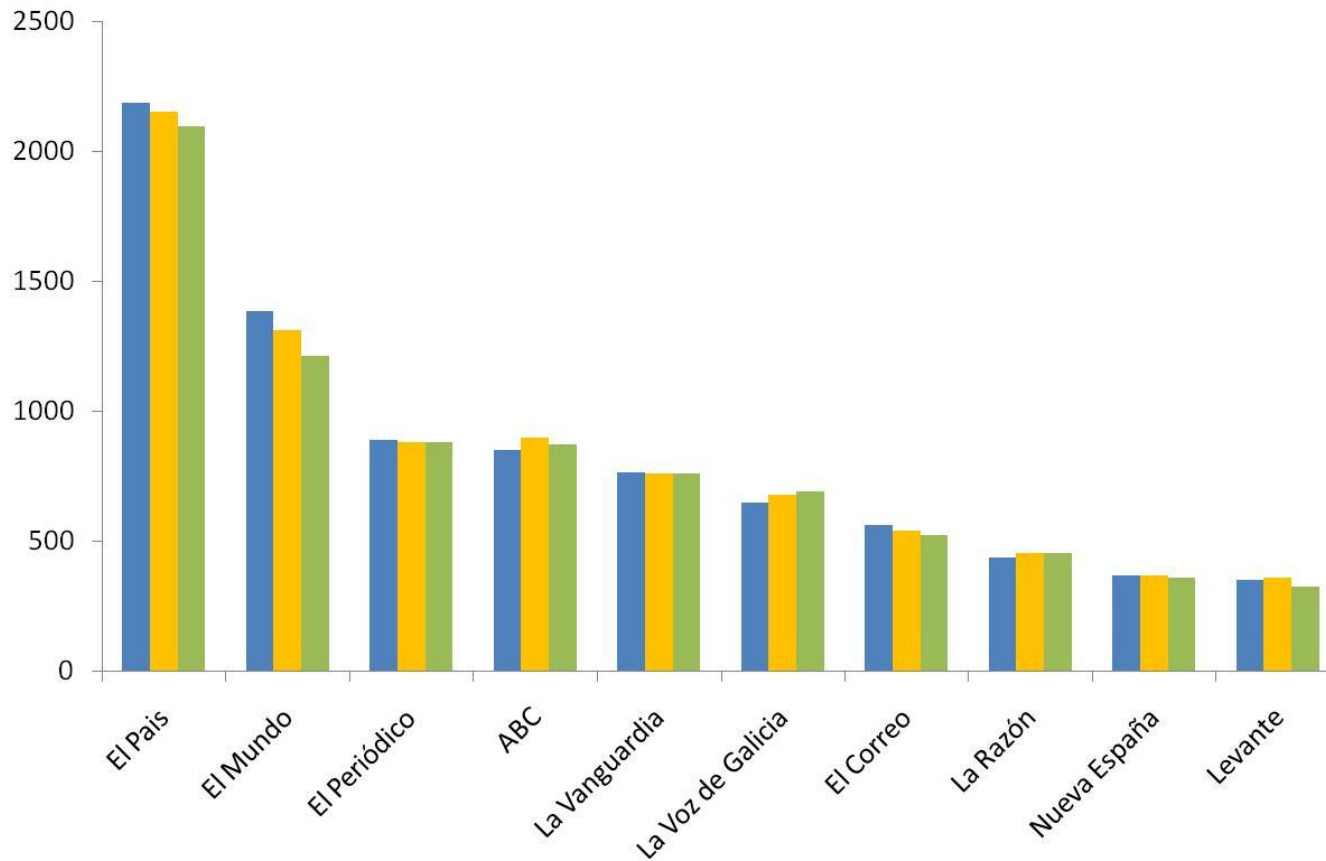
# La Estadística en los medios

- Estadística (del latín *status*: situación)
  - ▣ Fotos de la misma manifestación...



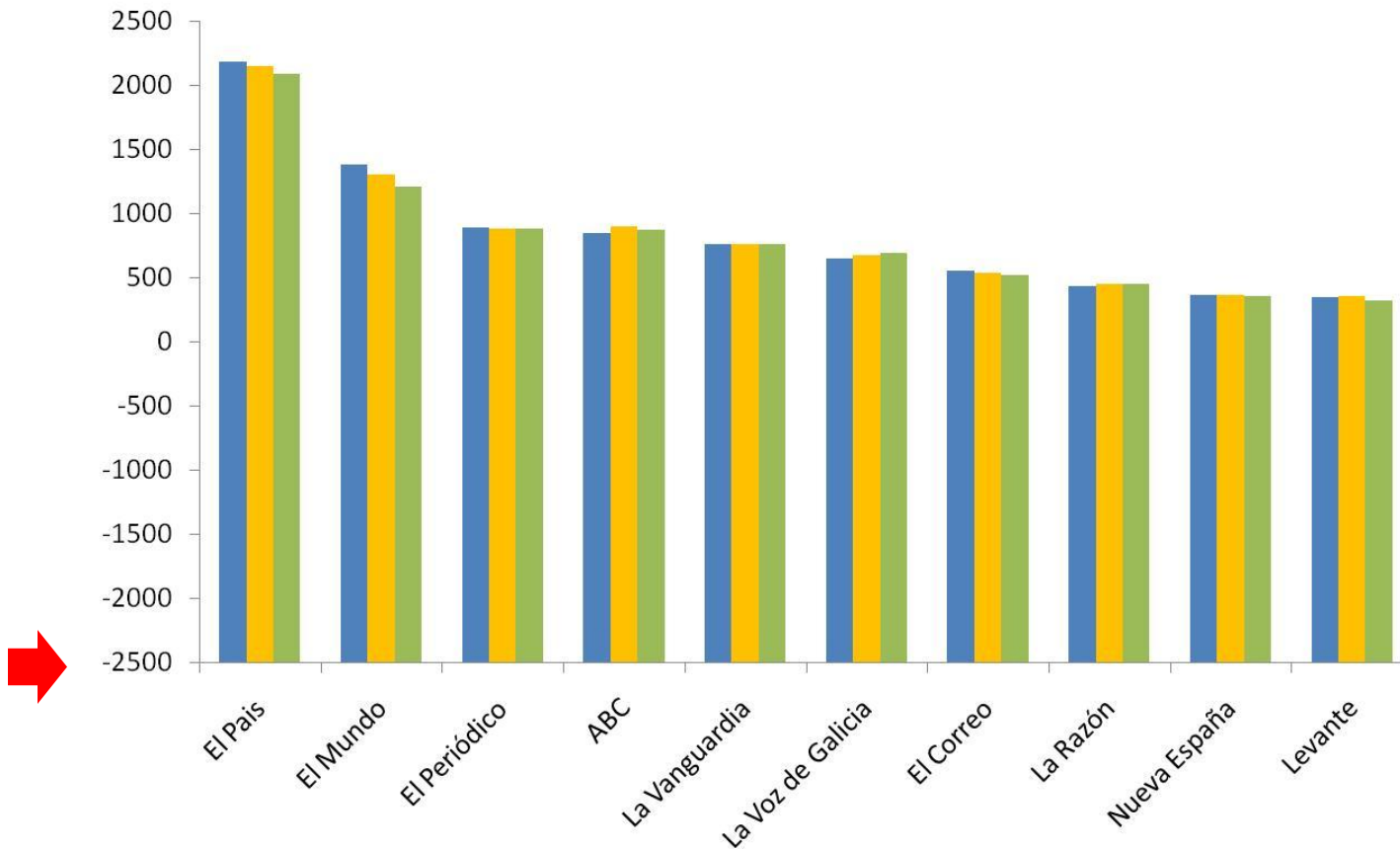
# La Estadística en los medios

## □ La verdad:



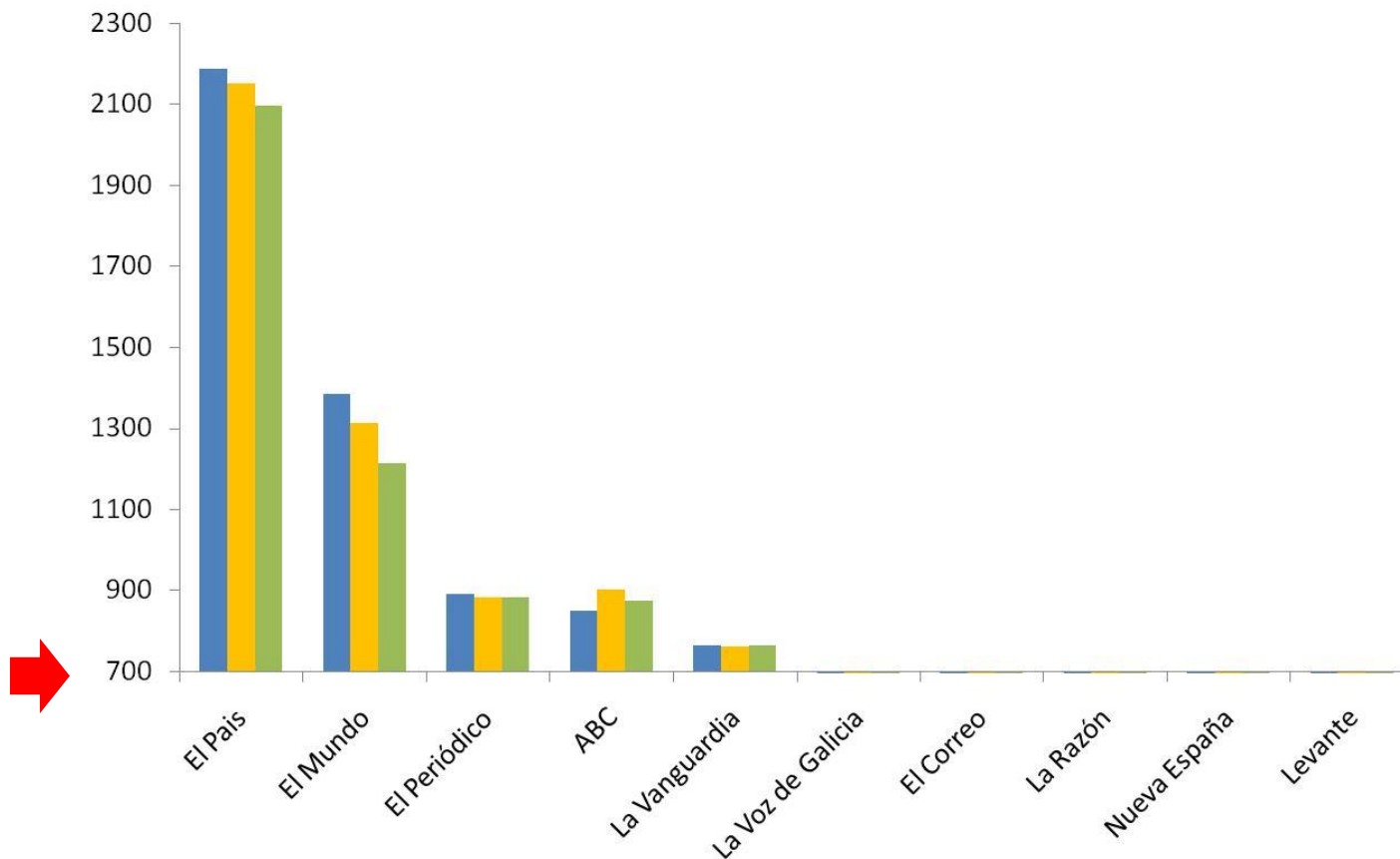
# La Estadística en los medios

## □ La de unos...



# La Estadística nos medios

## □ La de los otros...



# Estadística en los medios

## □ Estudios de Inserción Laboral



# Breve historia de la Estadística

- Censos:
  - ▣ Desde los Egipcios...
- Siglos XV y XVI
  - ▣ Nacimiento de la Demografía
  - ▣ Graunt: predicciones sobre evolución de la población
- Siglos XVII y XVIII
  - ▣ Cálculo de probabilidades
  - ▣ Jacques Quételet y Christiaan Huygens : aplicación de la Estadística en las Ciencias Sociales
- Siglos XIX y XX
  - ▣ Francis Galton: colabora con Mendel y Darwin en el nacimiento de la genética. Concepto de Regresión.
  - ▣ Laplace, Gauss, Bernouilli,
  - ▣ Pearson, Fisher, Tukey, Kolmogorov,.....

# ¿Qué es la Estadística Descriptiva?

## Estadística descriptiva

- Técnicas para describir y analizar un grupo dado de datos, sin extraer conclusiones (inferencias) sobre la población a la que pertenecen.

## Inferencia Estadística

- Es la parte de la Estadística que trata las condiciones bajo las cuales las inferencias extraídas a partir de una muestra son válidas

# ¿Qué es la Estadística descriptiva?

- La **estadística descriptiva** es una parte de la estadística que se dedica a analizar y representar los datos. Este análisis es muy básico, pero **fundamental en todo estudio...**

*Wikipedia*

# Población, muestra, variable



## Población

- Colectivo de individuos sobre los que se desea extraer alguna conclusión.



## muestra

- Subconjunto de la Población...



## Variable

- Cada una de las características medidas sobre los individuos de una muestra

# Población, muestra, variable



# Tablas de frecuencias

## □ Tablas de frecuencias:

- ▣ Para representar y resumir la información contenida en una muestra de tamaño  $n$
- ▣ modalidades: cada un de los valores que puede tomar una variable (cualitativa o cuantitativa discreta)

$$c_i, i = 1, \dots, k$$

- ▣ Intervalos de clase: en el caso de variables cuantitativas continuas, se agrupan los distintos valores obtenidos en la muestra en intervalos.

# Tablas de frecuencias

□ Para cada modalidad, definimos:

▣ Frecuencia absoluta: número de individuos en la modalidad:

$$n_i$$

▣ Frecuencia relativa: proporción de individuos en la modalidad:

$$f_i = \frac{n_i}{n}$$

▣ Frecuencia absoluta acumulada:  $N_i = n_1 + \dots + n_i$

▣ Frecuencia relativa acumulada:  $F_i = f_1 + \dots + f_i$

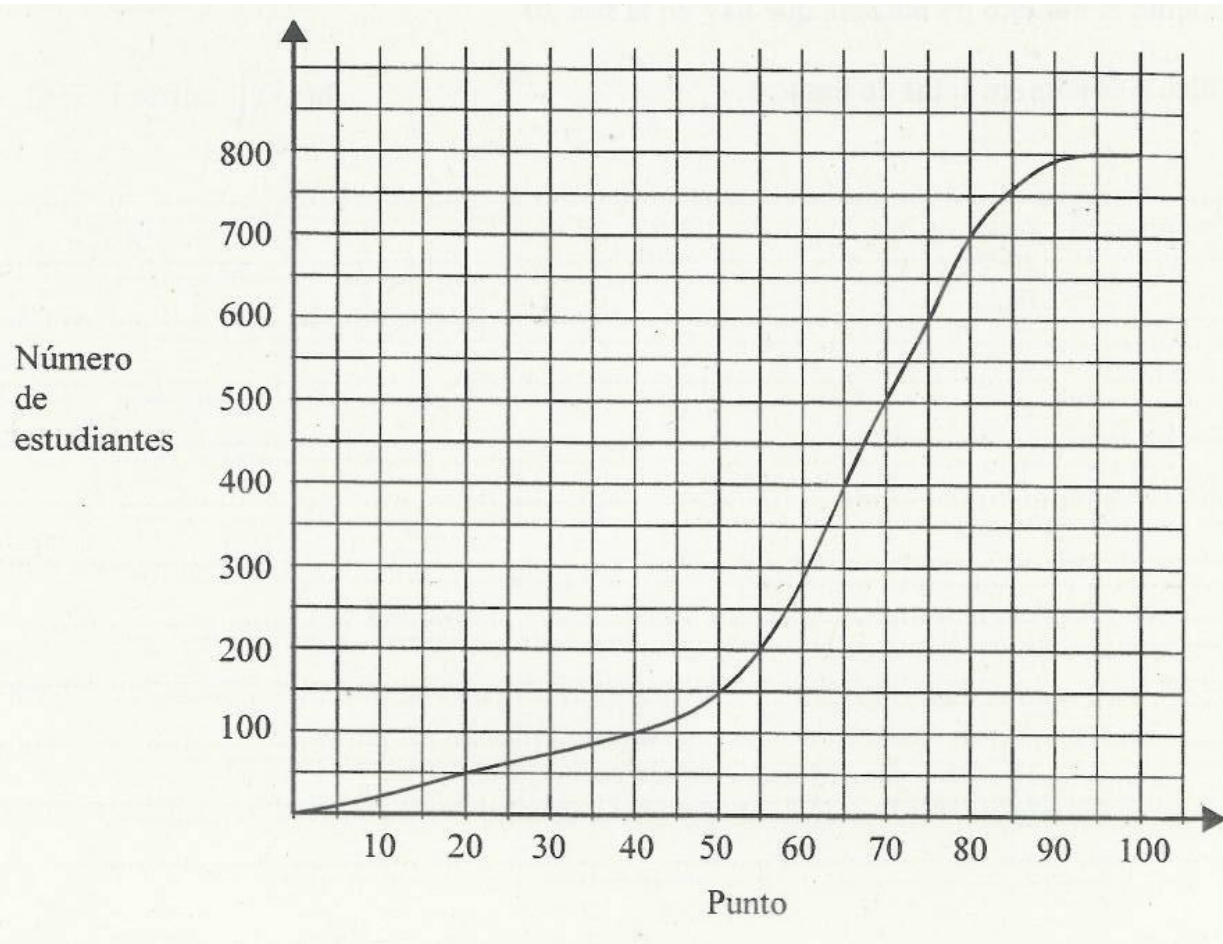
# Tablas de frecuencias

modalidad	Frecuencia absoluta	Frecuencia relativa	F. Absoluta acumulada	F. Relativa acumulada
$c_1$	$n_1$	$f_1$	$N_1$	$F_1$
$c_2$	$n_2$	$f_2$	$N_2$	$F_2$
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$
$c_k$	$n_k$	$f_k$	$N_k = n$	$F_k = 1$
TOTAL	$n$	$1$		

# Tablas de frecuencias

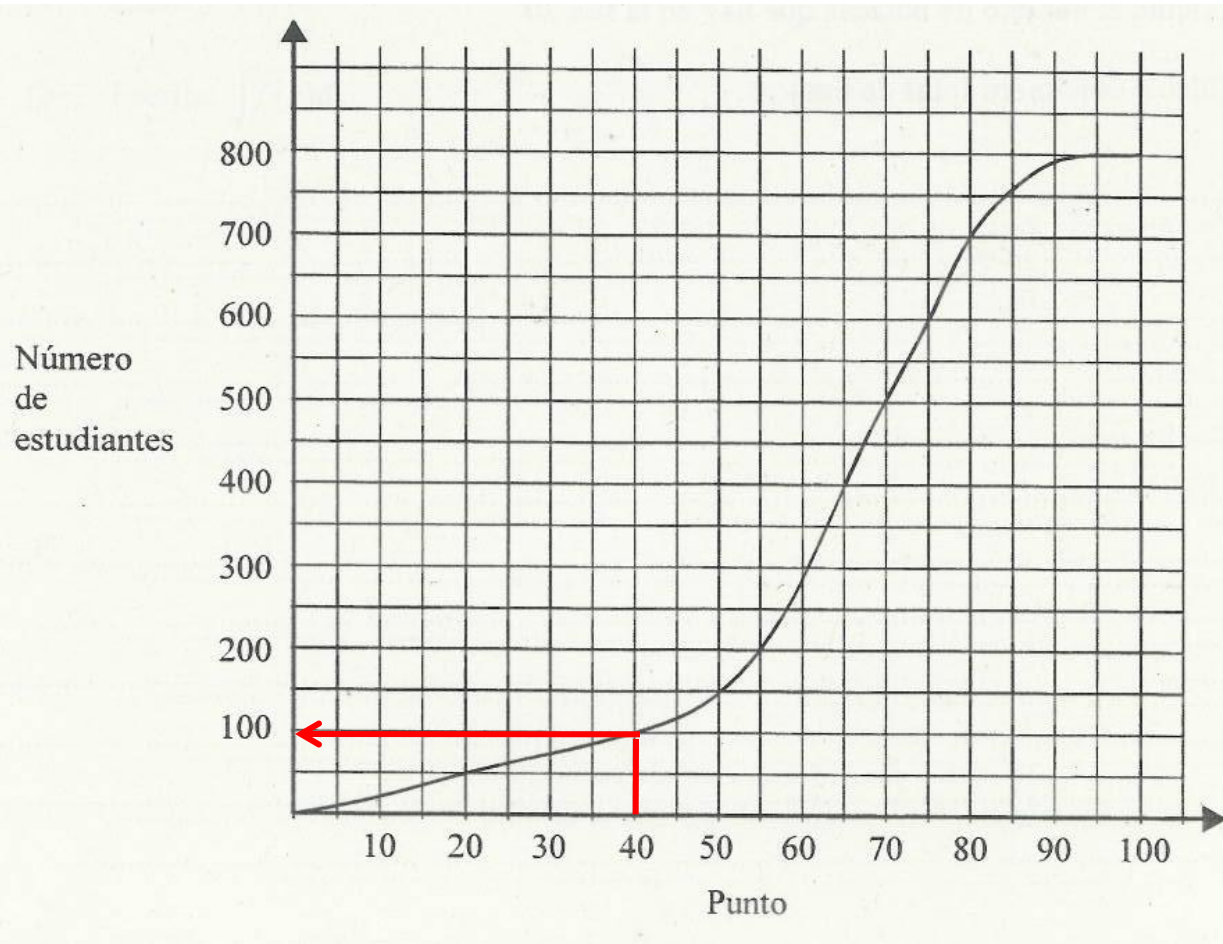
- Construcción de intervalos de clase:
  - ▣ Cada intervalo se considerará como una modalidad
  - ▣ Se elige un representante (marca de clase) de cada intervalo: o punto medio.
  - ▣ Cuántos? Entero más próximo a  $\sqrt{n}$
  - ▣ Todos iguales? Normalmente, si.
  - ▣ Como escogemos el primero e el último punto? (menor que el mínimo y mayor que el máximo de la muestra)
- Todo depende de la *inferencia*...
  - ▣ Frecuencias relativas: probabilidad!
  - ▣ Frecuencias relativas acumuladas: distribución!

# Frecuencias acumuladas



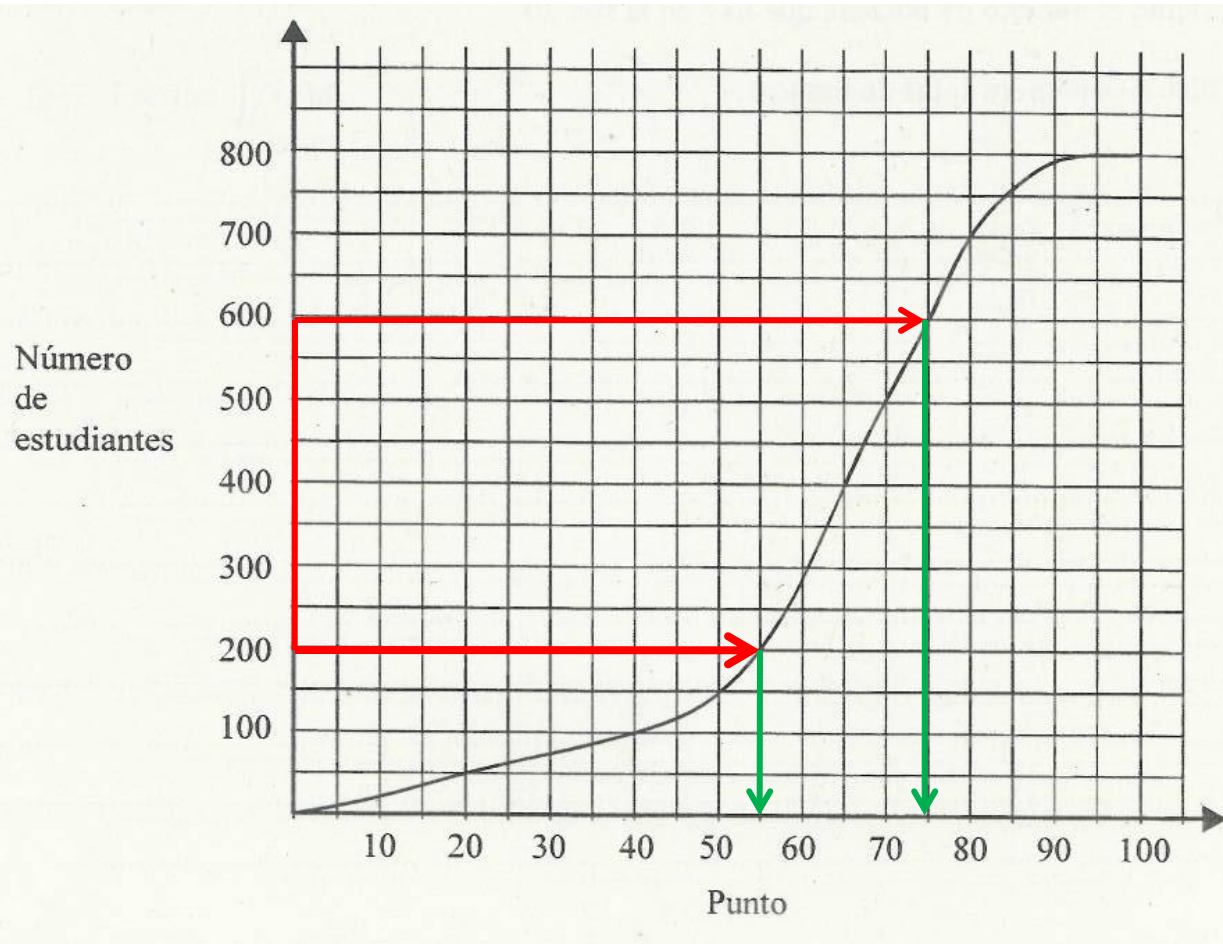
1. Cuántos obtuvieron 40 puntos o menos en el examen?
2. Entre que das notas se ubica el 50% central de los resultados?

# Frecuencias acumuladas



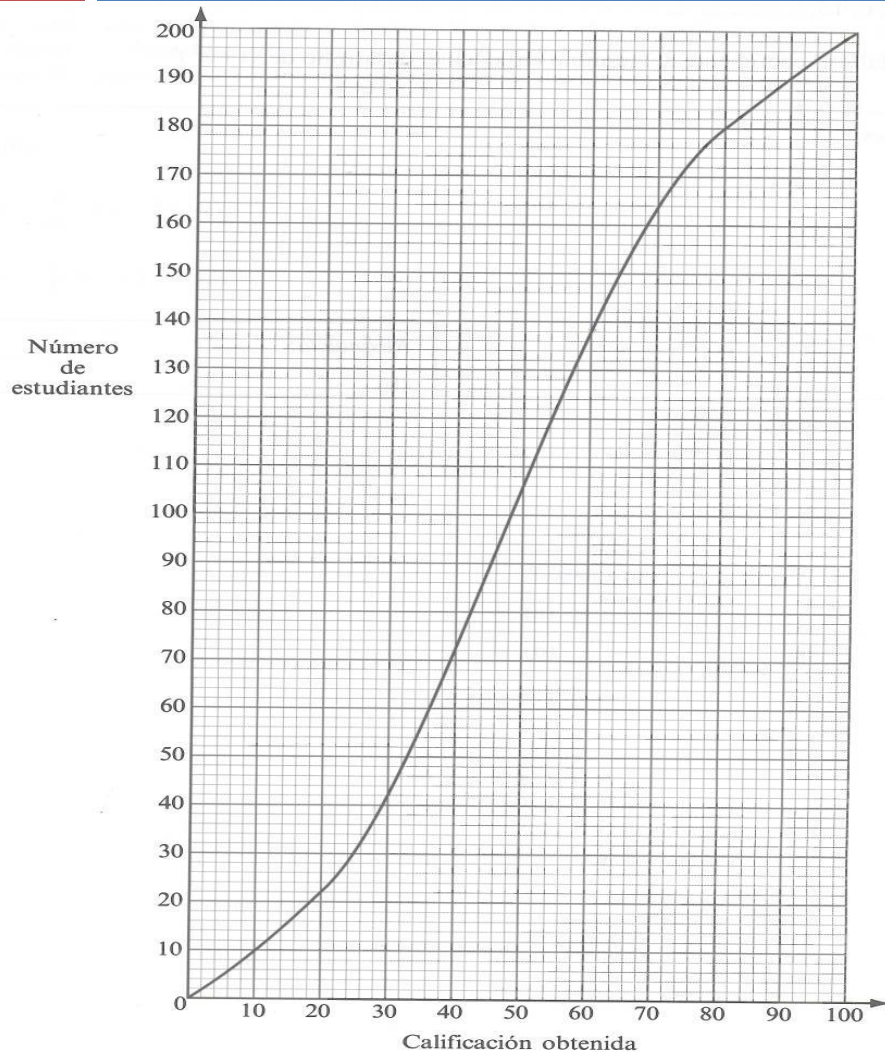
1. Cuántos obtuvieron 40 puntos o menos en el examen?
2. Entre que dos notas se ubica el 50% central de los resultados?

# Frecuencias acumuladas



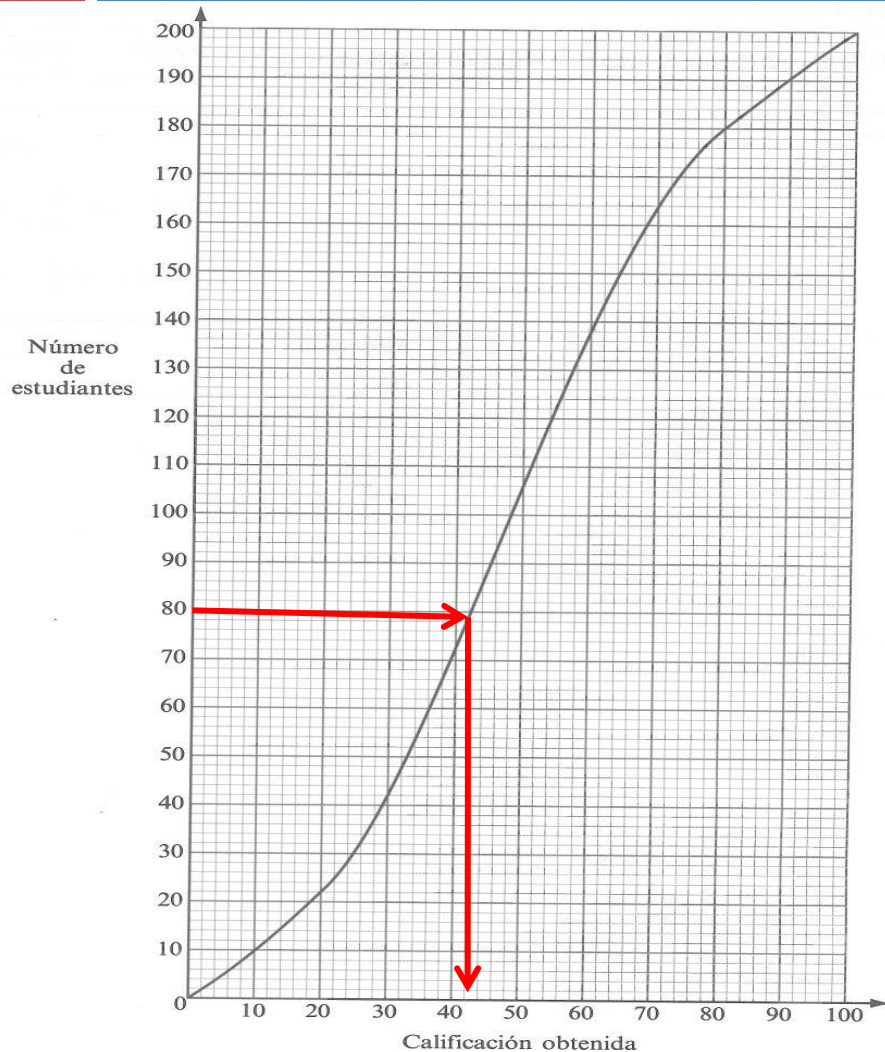
1. Cuántos obtuvieron 40 puntos o menos en el examen?
2. Entre que dos notas se ubica el 50% central de los resultados?

# Frecuencias acumuladas



1. El 40% no supera el examen...  
cuál es la calificación mínima para aprobar?

# Frecuencias acumuladas



1. El 40% no supera el examen... cuál es la calificación mínima para aprobar?

# Representaciones gráficas

**Métodos gráficos para variables  
cualitativas**

**Diagrama de barras**

**Diagrama de sectores**

# Representaciones gráficas

## □ Diagrama de barras

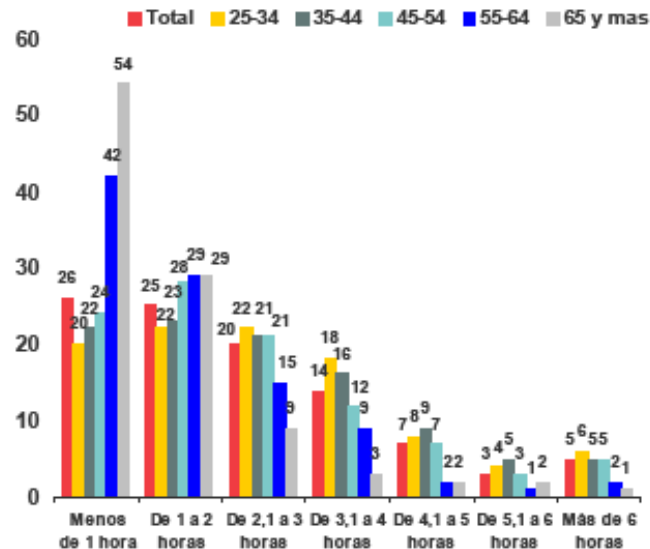
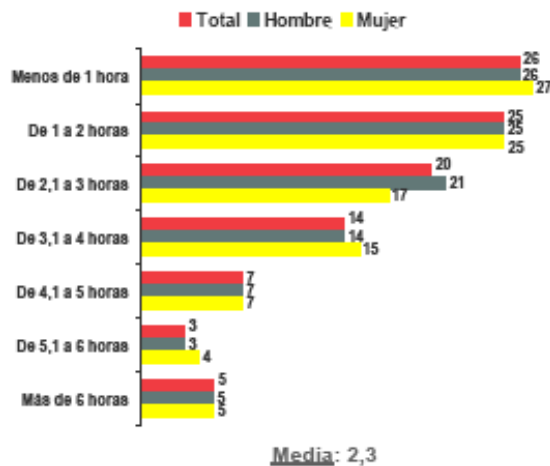
### TIEMPO DEDICADO A LAS RELACIONES DE CAMA/SEXUALES A LA SEMANA

-SEXO Y EDAD-



En general, 1 de cada 2 entrevistados dedican menos de 2 horas a las relaciones de cama/sexuales a la semana. No apreciándose diferencias entre los hombres y las mujeres.

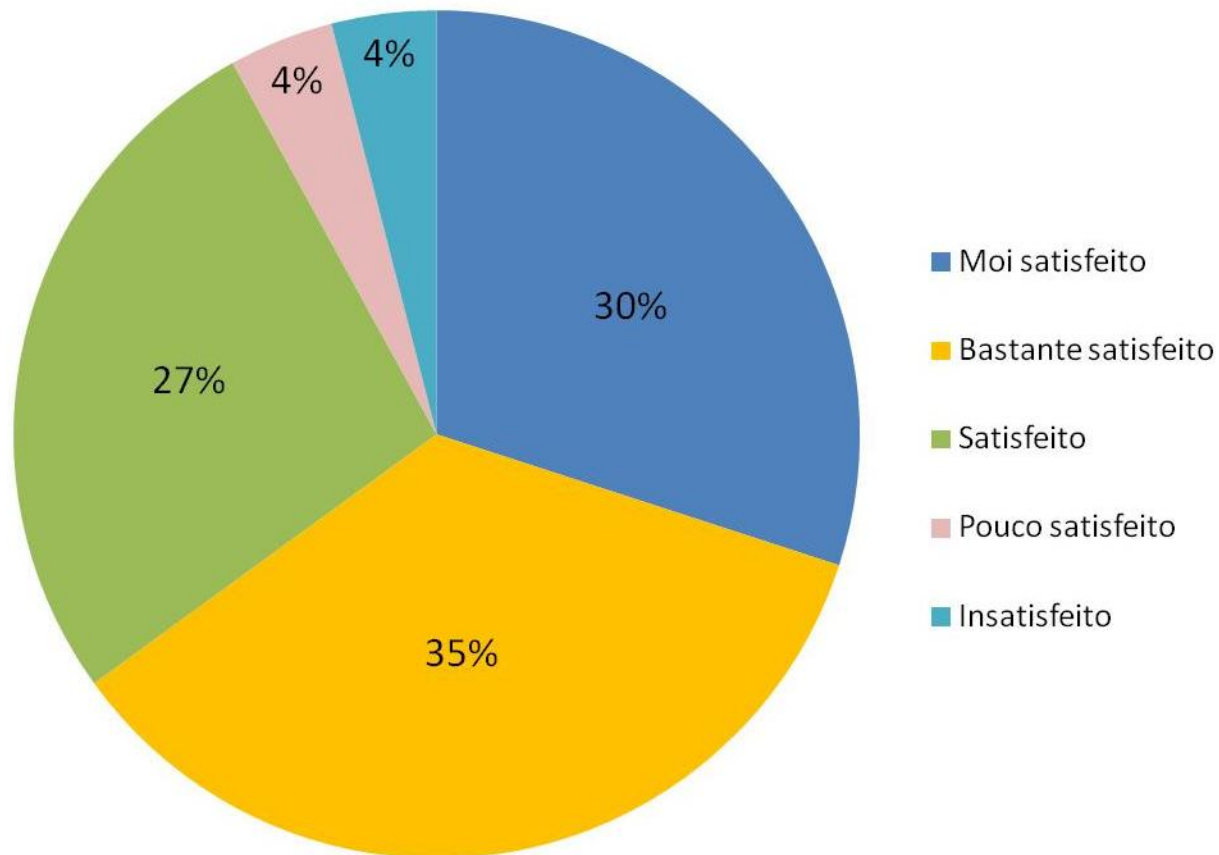
Uno de cada 2 entrevistados mayores de 64 años dedica menos de 1 hora a la semana a sus relaciones de cama/sexuales.



P.3b. ¿Cuánto tiempo dedica a la semana a las relaciones de cama/sexuales?

# Representaciones gráficas

## □ Diagrama de sectores



# Representaciones gráficas

## Métodos gráficos para variables cuantitativas

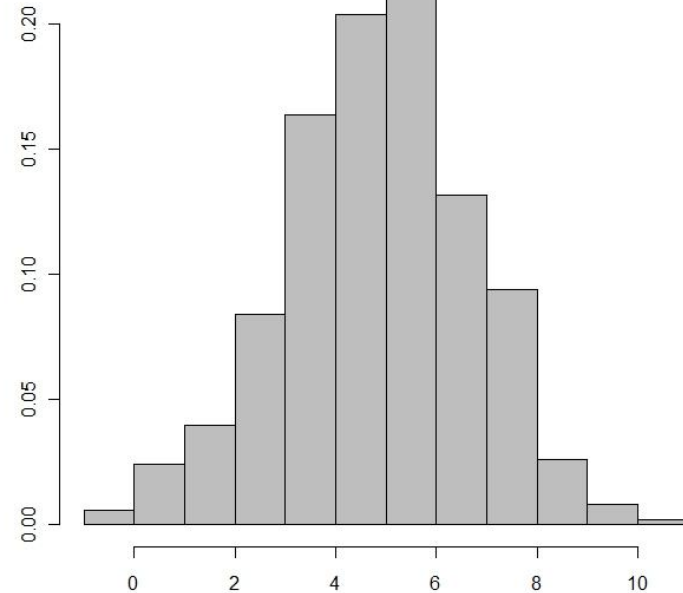
Histograma

Diagrama de tallo y hojas

Diagrama de caja (Box-Plot) \*

# Representaciones gráficas

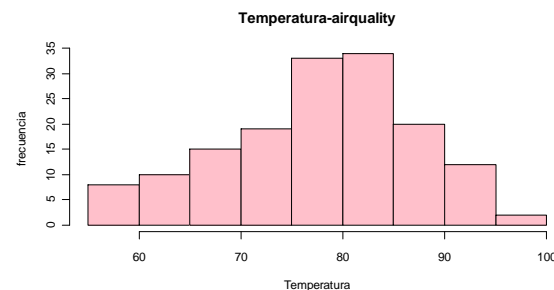
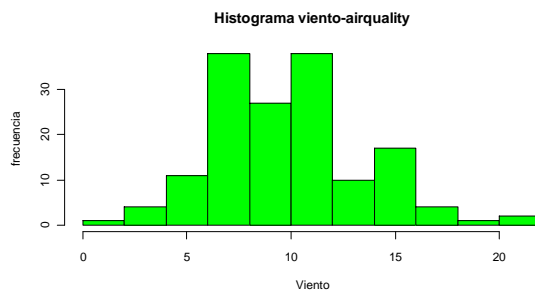
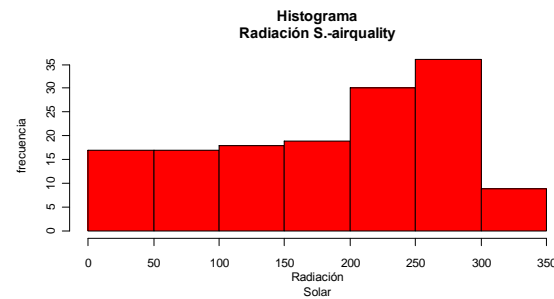
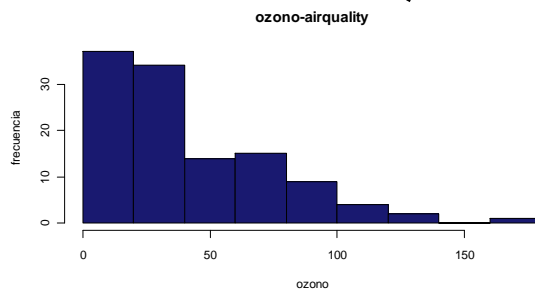
- HISTOGRAMA: Cómo se construye?
  - ▣ El histograma es una generalización del diagrama de barras para el caso de variables cuantitativas continuas.
  - ▣ Sobre cada intervalo de clase, Se levanta un rectángulo de altura la frecuencia (absoluta o relativa) del intervalo (en caso de que los intervalos tengan distinto tamaño se hacen las áreas del rectángulo proporcionales a dicha frecuencia, calculando las alturas).



# Representaciones gráficas

## □ HISTOGRAMA: Qué información contiene?

- ▣ La forma *general* de la distribución.
- ▣ Simetría o asimetría.
- ▣ Unimodalidad o multimodalidad.
- ▣ Sobre las frecuencias relativas: intuición sobre la función de densidad (polígono de frecuencias)



# Representaciones gráficas

- HISTOGRAMA: Cuántos intervalos?
  - Número de intervalos: el entero más próximo a  $\sqrt{n}$  (normalmente entre 5 y 20).
  - ○ histograma é sensible á selección do número de intervalos:
    - Si tenemos una distribución bimodal, y escogemos intervalos muy amplios... no se detecta!
    - Si escogemos intervalos muy pequeños, pueden aparecer características en el histograma que sean sólo debidas a ruído.

# Representaciones gráficas

## □ DIAGRAMA DE TALLO Y HOJAS (John Tukey)

- Consideremos los siguientes datos, en orden ascendente:

- 8, 13, 16, 25, 26, 29, 30, 32, 37, 38, 40, 41, 44, 47, 49, 51, 54, 55, 58, 61, 63, 67, 75, 78, 82, 86, 95

- El diagrama se construye colocando en una primera columna los primeros dígitos, y los demás en las siguientes columnas...

0		8
1		36
2		569
3		0278
4		01479
5		1458
6		137
7		58
8		26
9		5



# Medidas características

- Es una cantidad numérica calculada sobre una Población

Parámetro



- Es una cantidad numérica calculada sobre una muestra

Estadístico



# Medidas características

Tendencia

Dispersión

Forma

Tendencia  
central

Tendencia no  
central

Varianza  
Desviación  
típica

Asimetría  
Curtosis

Media  
Mediana  
Moda

Cuartiles  
Deciles  
Percentiles

# Medidas características

## □ De tendencia central

□ Indican el punto alrededor del cual se sitúan los datos.

□ Media aritmética:

$$\bar{x} = \frac{1}{n} \sum_{i=1}^n x_i$$

□ Propiedades:

■ Entre el mínimo y el máximo:

$$\min\{x_1, \dots, x_n\} \leq \bar{x} \leq \max\{x_1, \dots, x_n\}$$

■ Linealidad

■ Media de las desviaciones con respecto a la media: 0

■ Media de los cuadrados con respecto a una constante: mínima para la media aritmética

# Medidas características

## □ De tendencia central

### □ Media aritmética:

$$\bar{x} = \frac{1}{n} \sum_{i=1}^n x_i$$

### □ Propiedades:

#### ■ No es robusta!

- Media truncada: eliminar un porcentaje de los datos más extremos.
- Media recortada: substituir un porcentaje de los datos más extremos por el valor del punto de corte.

# Medidas características

## □ De tendencia central

### □ Media aritmética:

$$\bar{x} = \frac{1}{n} \sum_{i=1}^n x_i$$

### □ Sobre una Tabla de frecuencias:

$$\bar{x} = \frac{1}{n} \sum_{i=1}^k n_i c_i$$

# Medidas características

## □ De tendencia central

- Mediana: es el punto en el que encontramos el 50% de los datos de una muestra ordenada.
  - Intervalo mediano (la frecuencia absoluta en su extremo inferior es menor que  $1/2$ )

# Medidas características

## □ De tendencia central

- Moda: (variables discretas o cualitativas) es el valor que más se repite.
  - Intervalo modal: variables cuantitativas continuas (intervalo con mayor frecuencia)

# Medidas características

## □ De tendencia no central

□ Cuartiles: los cuartiles Q1, Q2 y Q3 dividen la muestra en cuatro partes iguales. Es decir, por debajo de Q1 tenemos el 25%, por debajo de Q2 tenemos el 50% y por debajo de Q3 el 75% de los datos.

■ Q2 es la mediana

# Medidas características

## □ De tendencia no central

- Deciles:  $D_1, \dots, D_9$  dividen la muestra en 10 partes iguales (intervalos del 10%).
- Percentiles:  $P_1, \dots, P_{99}$  dividen a muestra en 100 partes iguales (intervalos del 1%)
- Cuantiles: valores en tanto por uno (0.25, 0.89, ...)

# Medidas características

## □ Medidas de Dispersión

□ Indican la mayor o menor concentración de los datos con respecto a las medidas de centralización.

□ Varianza:

$$s^2 = \frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})^2$$

□ Desviación típica:

$$s = \sqrt{\frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})^2}$$

# Medidas características

## □ Medidas de Dispersión

□ Sobre una Tabla de frecuencias:

□ Varianza:

$$s^2 = \frac{1}{n} \sum_{i=1}^k n_i (c_i - \bar{x})^2$$

□ Desviación típica:

$$s = \sqrt{\frac{1}{n} \sum_{i=1}^k n_i (c_i - \bar{x})^2}$$

# Medidas características

## □ **Medidas de Dispersión**

### □ Varianza y desviación típica

### □ Propiedades:

- Valores no negativos
- No linealidad! Pero no afectada por cambios de localización.
- La varianza puede calcularse como:

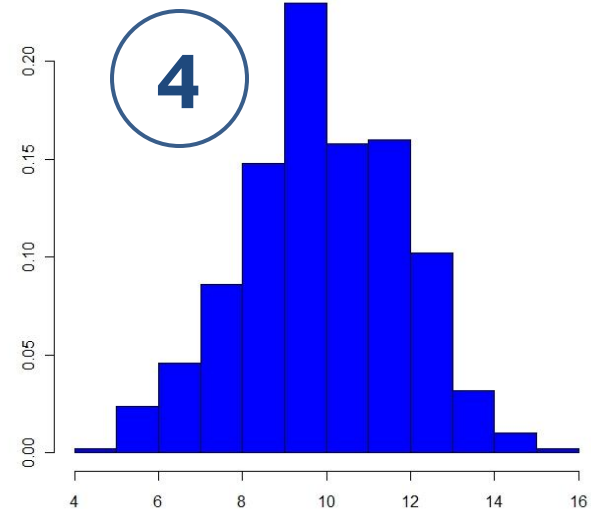
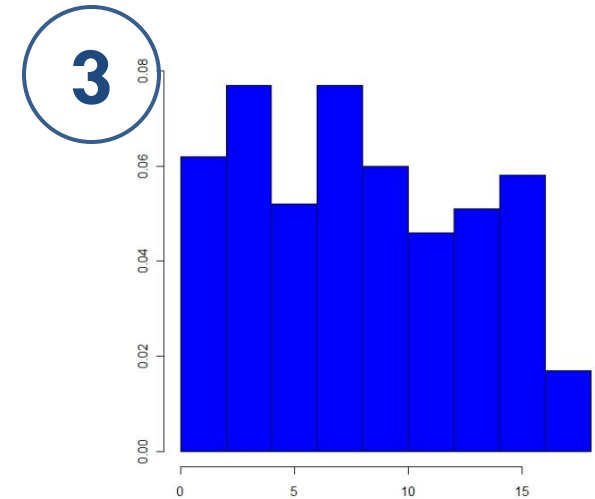
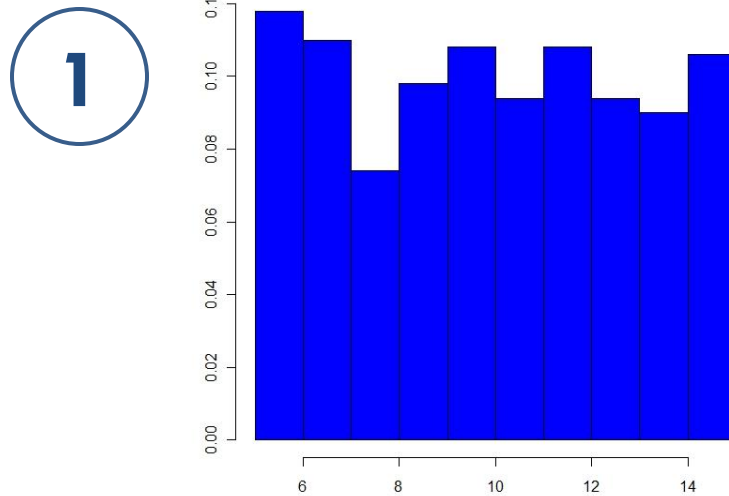
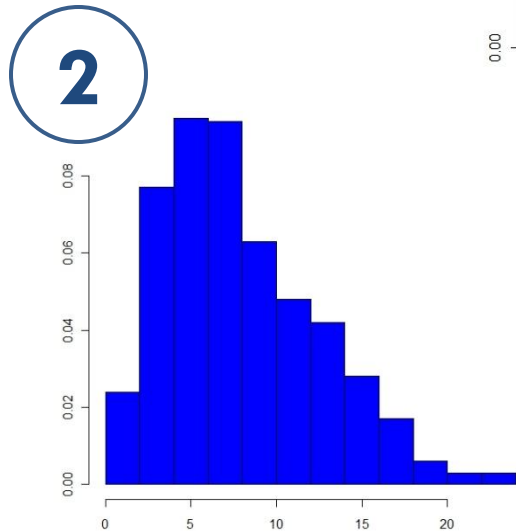
$$s^2 = \frac{1}{n} \sum_{i=1}^n x_i^2 - \bar{x}^2$$

# Medidas características

- **Medidas de Dispersión**
  - Rango muestral:  $\text{máx}(x) - \text{mín}(x)$
  - Rango intercuartílico:  $Q3 - Q1$

# Gráficas y medidas características

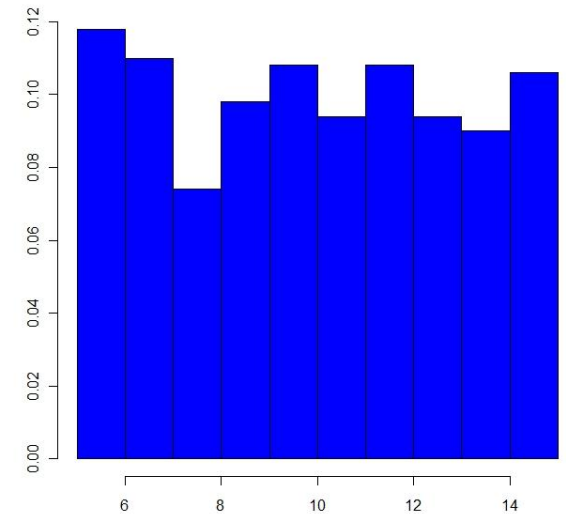
□ Un ejercicio con histogramas:



# Gráficas y medidas características

- Media
  - 9.93
- Desviación típica
  - 2.93
- Mediana
  - 9.88
- RIC
  - 5.11

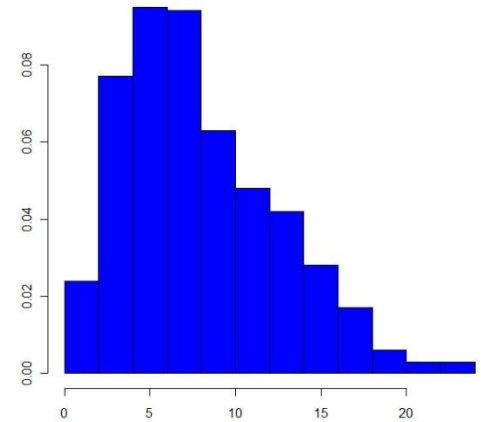
1



# Gráficas y medidas características

- Media
  - ▣ 7.98
- Desviación típica
  - ▣ 4.48
- Mediana
  - ▣ 7.29
- RIC
  - ▣ 6.22

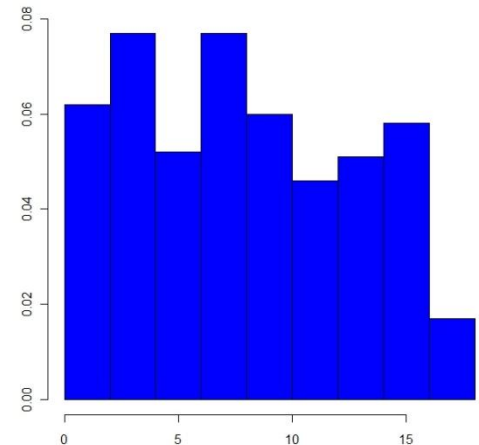
2



# Gráficas y medidas características

- Media
  - ▣ 7.88
- Desviación típica
  - ▣ 4.81
- Mediana
  - ▣ 7.46
- RIC
  - ▣ 8.31

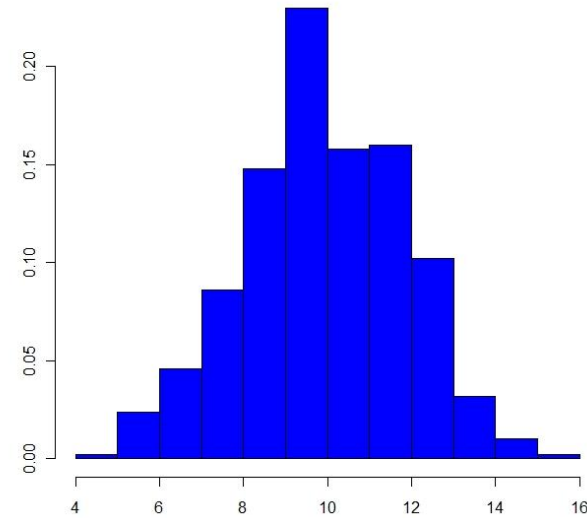
3



# Gráficas e medidas características

- Media
  - 9.90
- Desviación típica
  - 1.92
- Mediana
  - 9.84
- RIC
  - 2.65

4



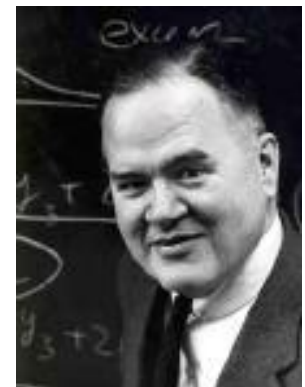
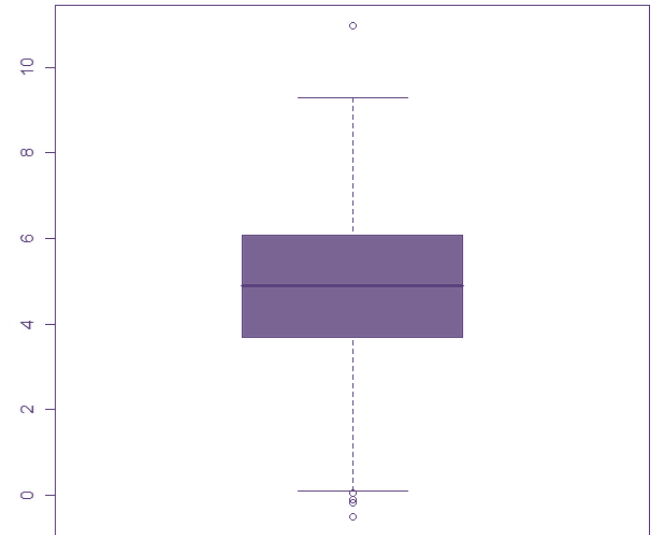
# Representaciones gráficas

- En 1977, John Tukey publica un método para representar gráficamente cinco estadísticos descriptivos.
- El diagrama de caja: **boxplot** o también “box and whisker plot”, que contiene las siguientes medidas:

- Mediana

- Primer y tercer cuartil

- Extremos(\*):  $LI = \max\{\min(x_i), Q_1 - 1.5(Q_3 - Q_1)\}$

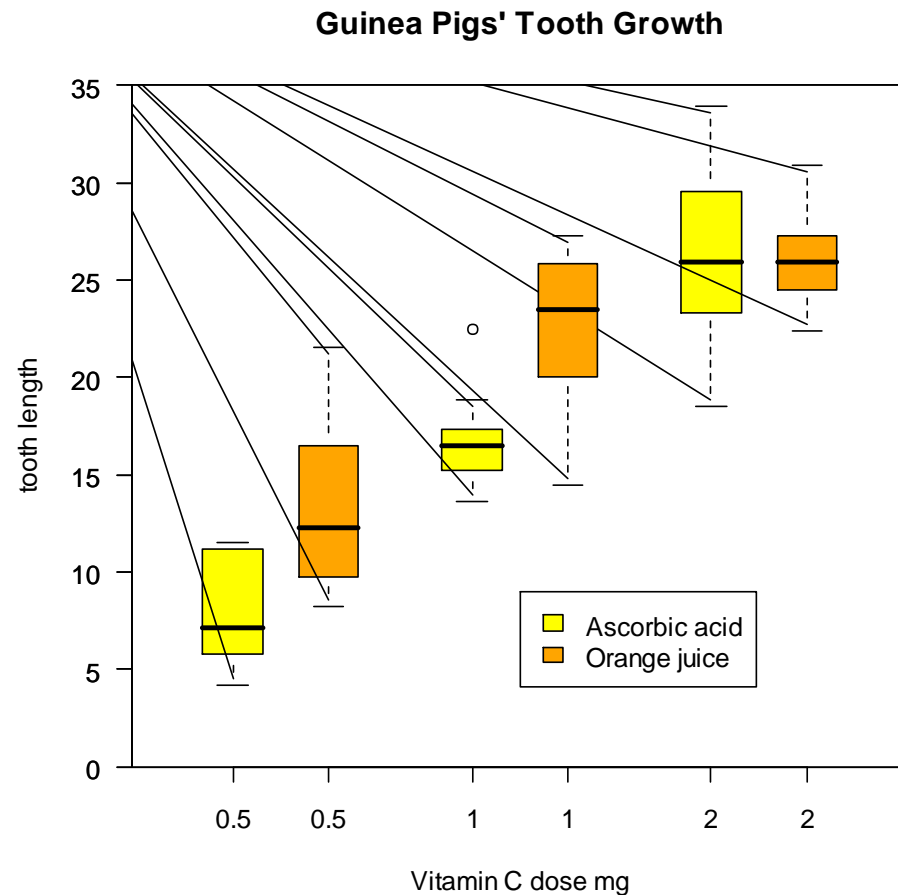
$$LS = \min\{\max(x_i), Q_3 + 1.5(Q_3 - Q_1)\}$$


John W. Tukey (1915-2000)

# Representaciones gráficas

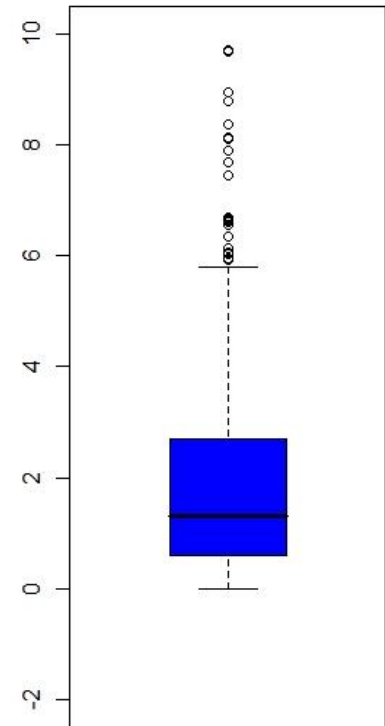
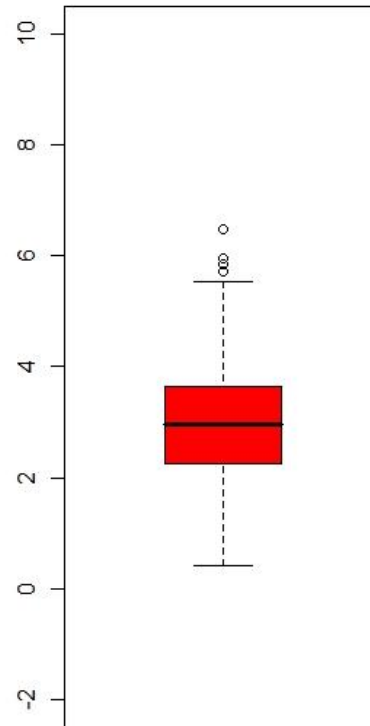
## □ BOXPLOT: Interpretación

- En la caja central, el Boxplot contiene el 50% de los datos. La línea interior marca la mediana, y los límites de la caja son el primer y tercer cuartil.
- Si la mediana no equidista de los vértices, entonces la distribución es asimétrica.
- Los puntos que quedan fuera del diagrama son datos atípicos (outliers).

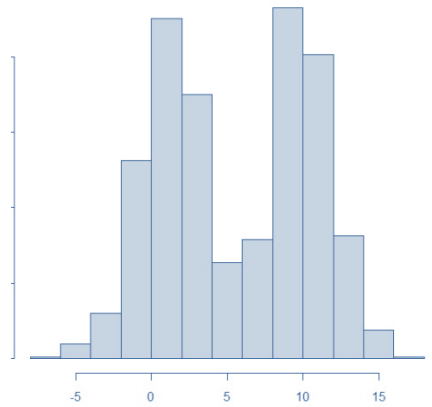
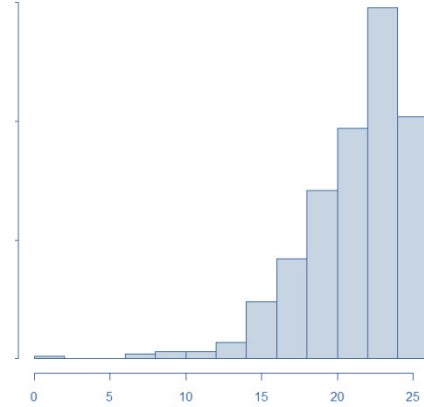
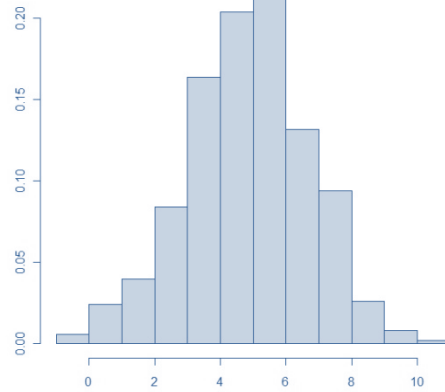
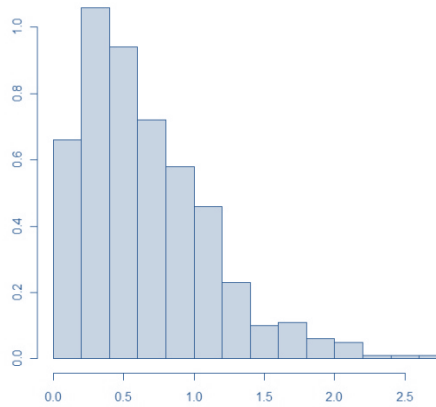


# Representaciones gráficas

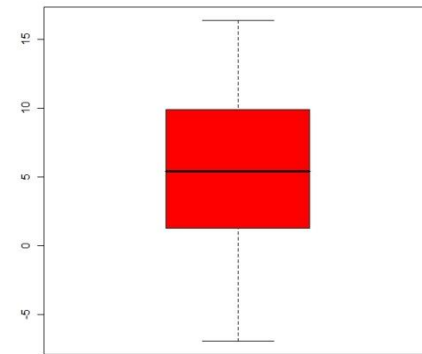
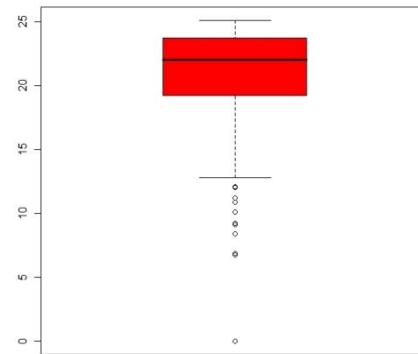
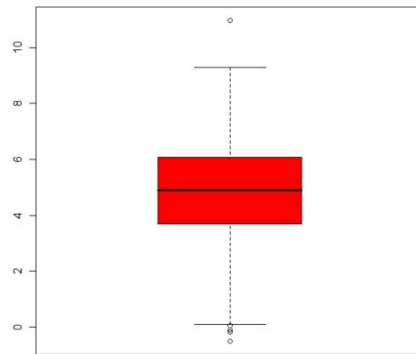
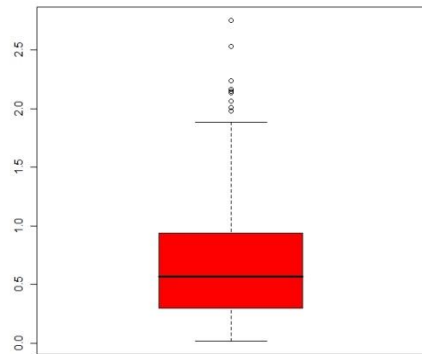
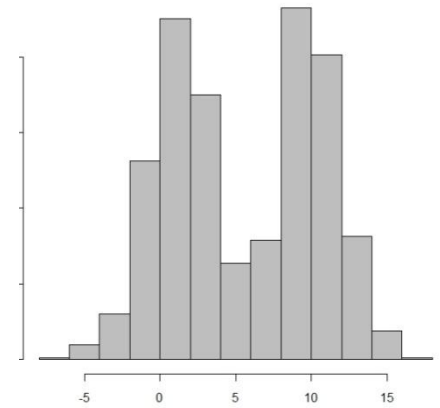
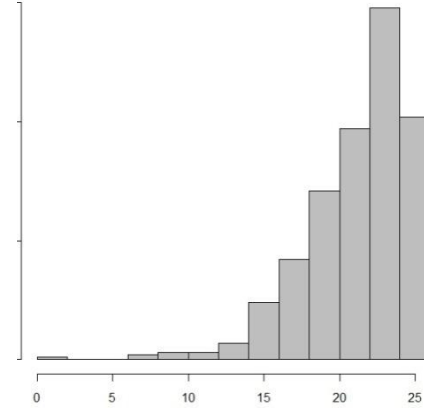
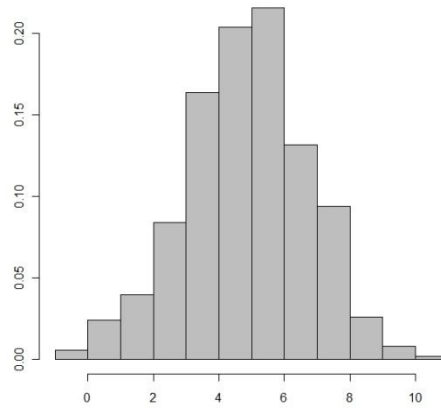
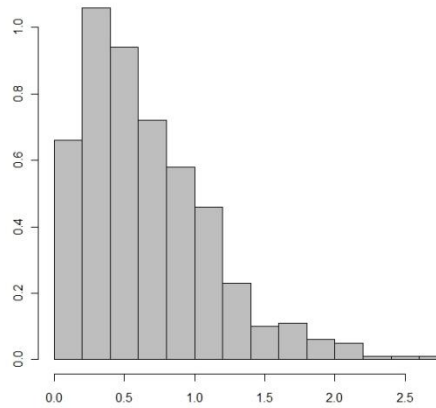
- **BOXPLOT: ventajas e inconvenientes**
  - ▣ Visualización de varios estadísticos al mismo tiempo!
  - ▣ Muestra los datos atípicos.
  - ▣ Útiles para comparar distribuciones *condicionadas* a variables categóricas.
  - ▣ ... pero no detectan multimodalidad!  
*Solución:* hacer un histograma y un boxplot



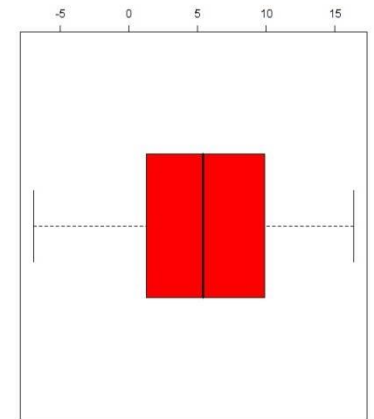
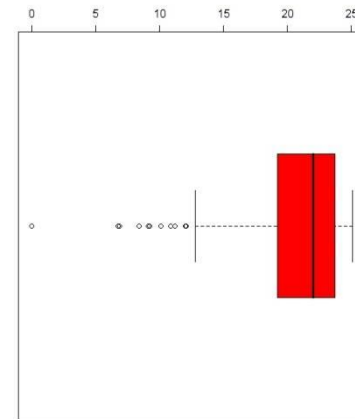
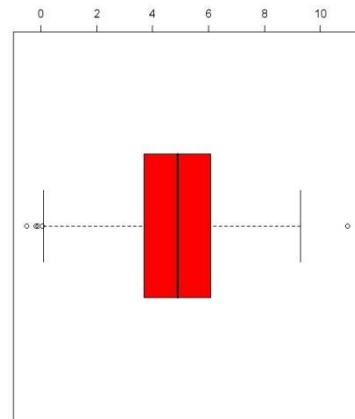
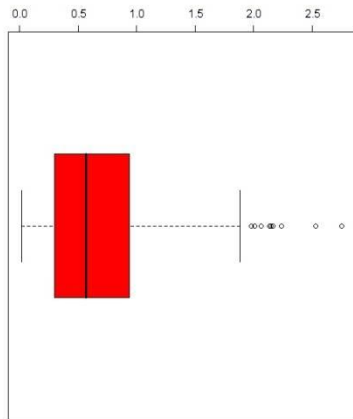
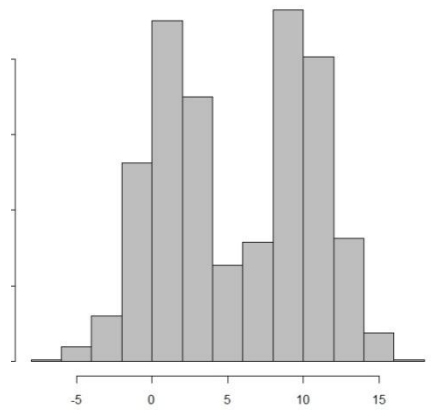
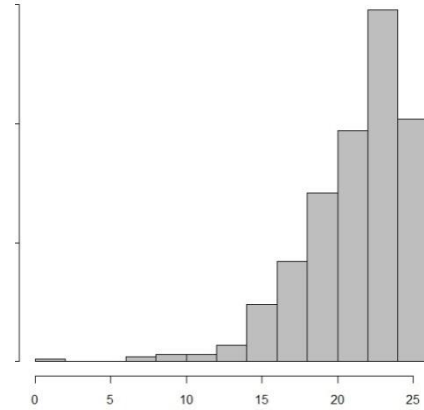
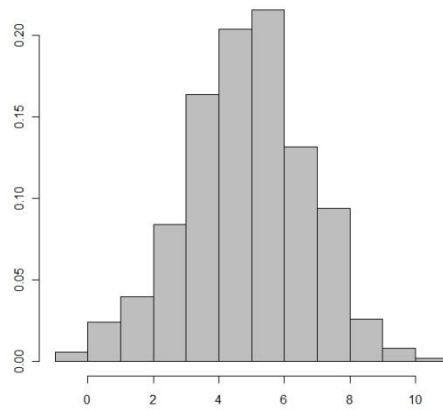
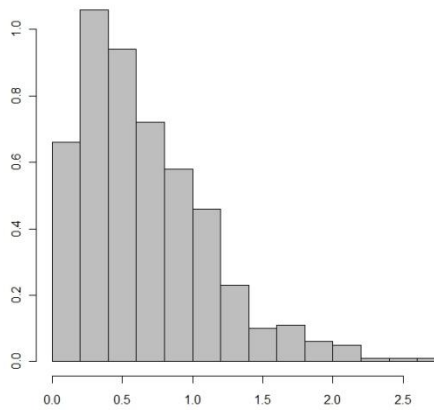
# Representaciones gráficas



# Representaciones gráficas



# Representaciones gráficas



# Medidas características

- Si tenemos dos conjuntos de datos con las siguientes desviaciones típicas:
  - ▣ Desv. Típica X: 2.005975
  - ▣ Desv. Típica Y: 309.0946

Cuál de los conjuntos es más *variable*?

# Medidas características

- Y si las medias son:
  - ▣ Media X: 0.99661
  - ▣ Media Y: 150.4486

Seguimos pensando lo mismo...?

# Medidas características

- Todas estas medidas dependen de las unidades de los datos

- Coeficiente de variación:  $CV = \frac{s}{\bar{x}}$

- Coef. X: 2.012799

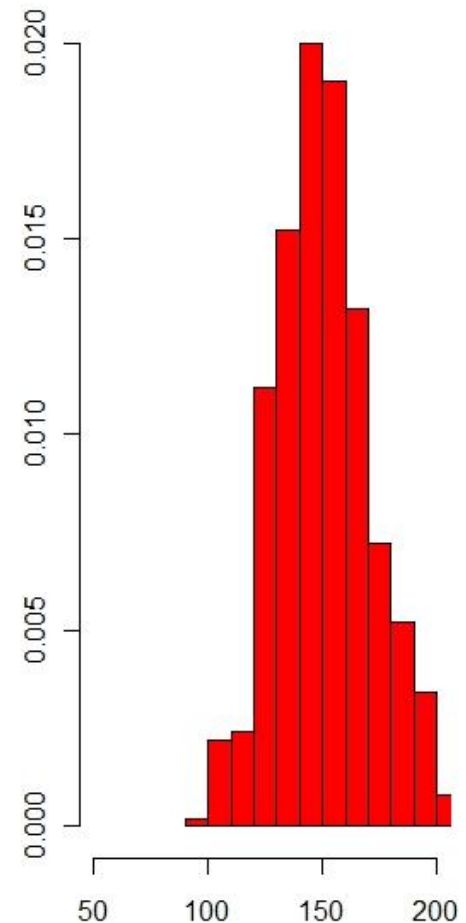
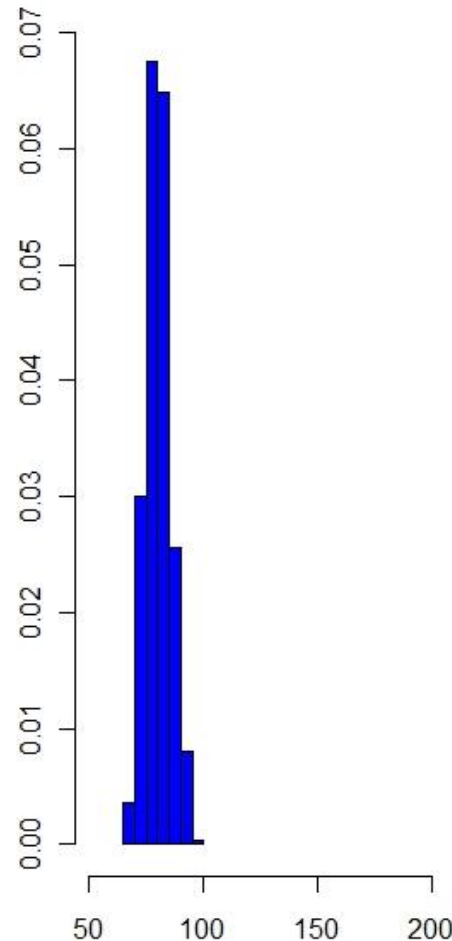
- Coef. Y: 2.054486

# Medidas características

- Para comparar muestras, debemos TIPIFICAR los datos:

$$z_i = \frac{x_i - \bar{x}}{s_x}$$

$$w_i = \frac{y_i - \bar{y}}{s_y}$$

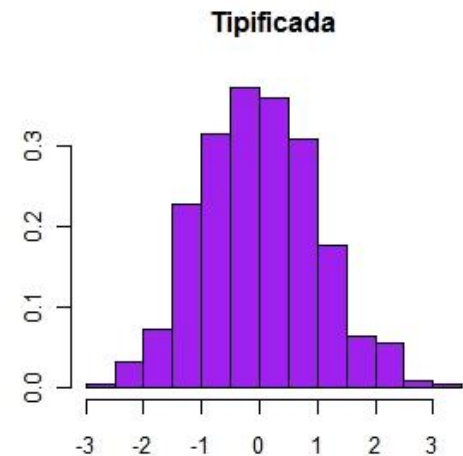
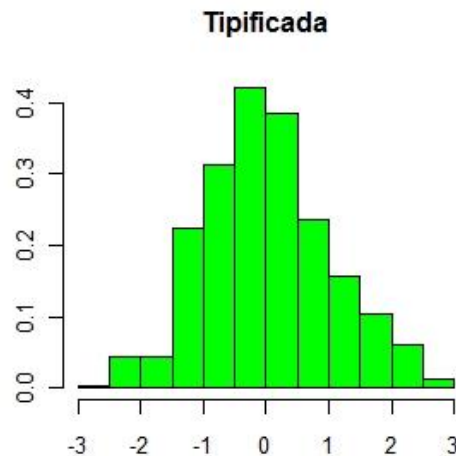
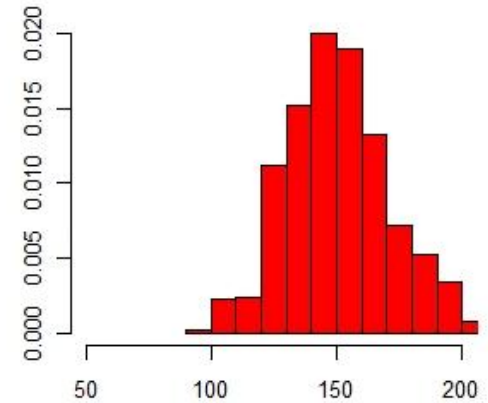
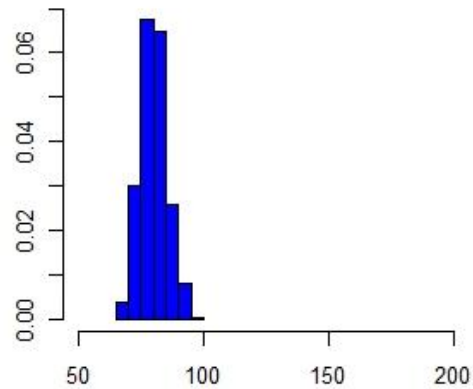


# Medidas características

- Para comparar muestras, debemos TIPIFICAR los datos:

$$z_i = \frac{x_i - \bar{x}}{s_x}$$

$$w_i = \frac{y_i - \bar{y}}{s_y}$$

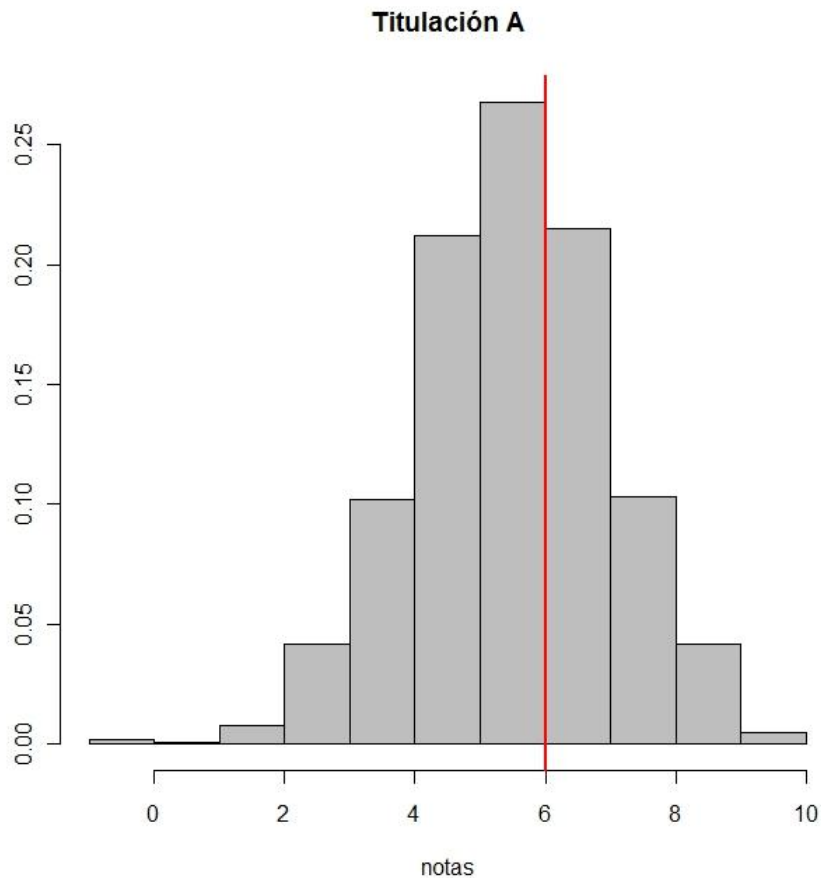


# Medidas características

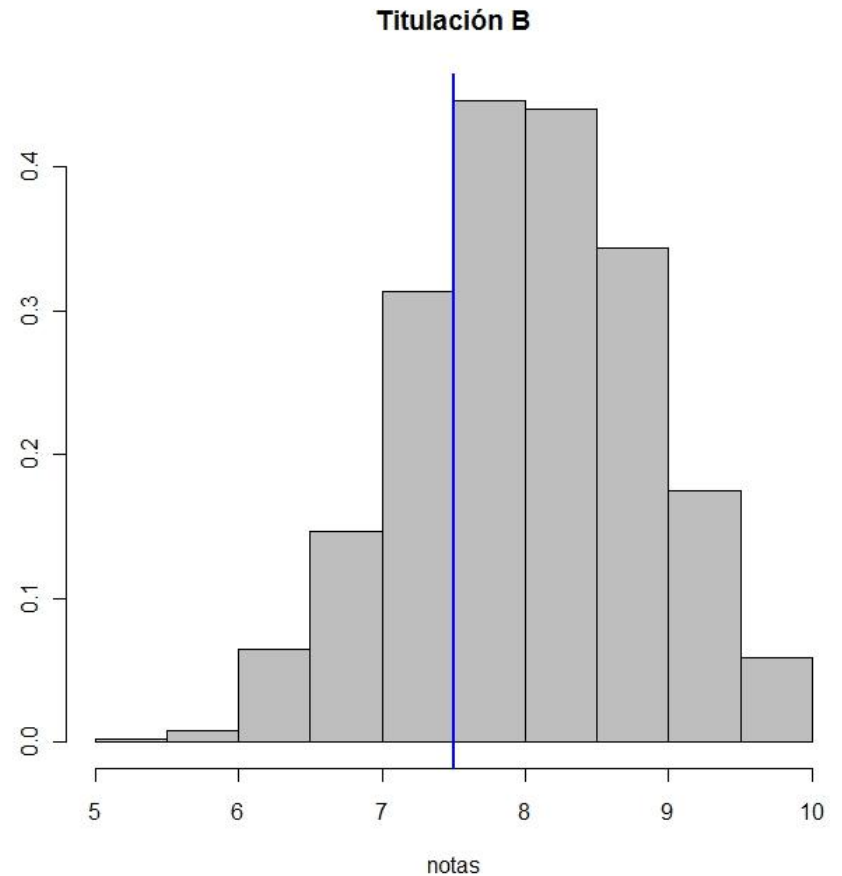
- Para la asignación de becas para estudios de posgrado, se consideran las notas medias del expediente:
  - ▣ Alumno titulación A: nota media de 6
  - ▣ Alumno titulación B: nota media de 7.5
    - A quién se le da la beca?

# Medidas características

□ Alumno titulación A:



□ Alumno titulación B:



# Medidas características

- Titulación A:  $media=5.5$ ,  $varianza=1.48$ 
  - En relación a su titulación, el alumno A (nota media de 6) es mejor que la media... pero hay que tener en cuenta la varianza para saber cuánto mejor.
- Titulación B:  $media=7.92$ ,  $varianza=0.56$ 
  - En relación a su titulación, el alumno B (nota media de 7.5) es peor que la media... pero también hay que tener en cuenta la dispersión.

# Medidas características

- Tipificamos: a cada uno de los valores de las dos titulaciones les restamos su media y los dividimos por su desviación típica:

- El valor 6 del alumno A queda tipificado como:

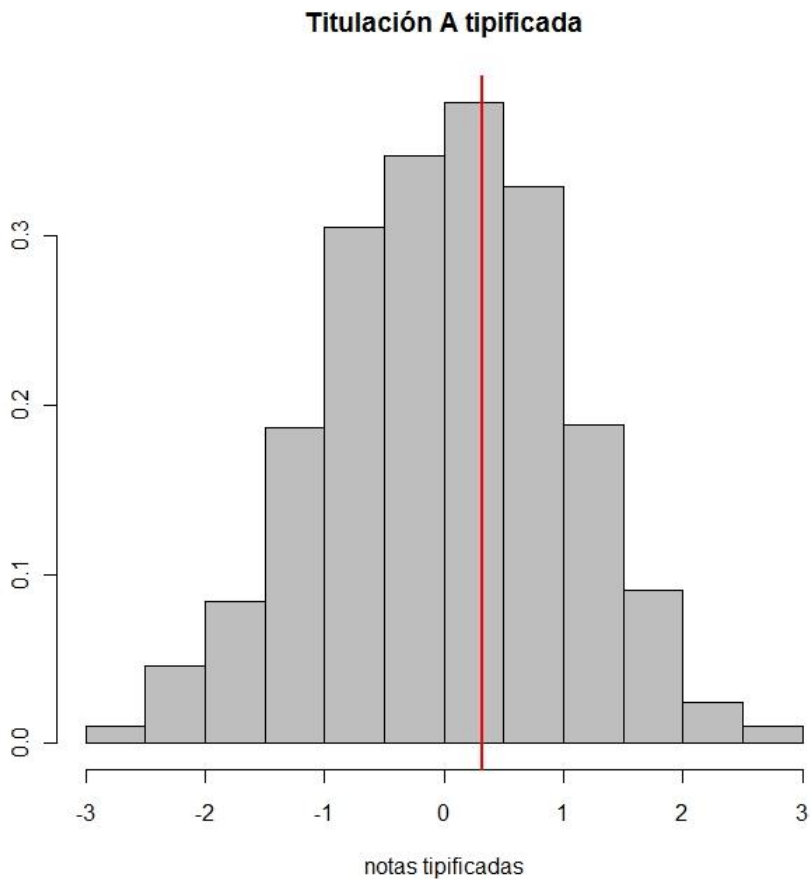
$$\frac{6 - 5.5}{1.48} = 0.32$$

- O valor 7.5 del alumno B queda tipificado como :

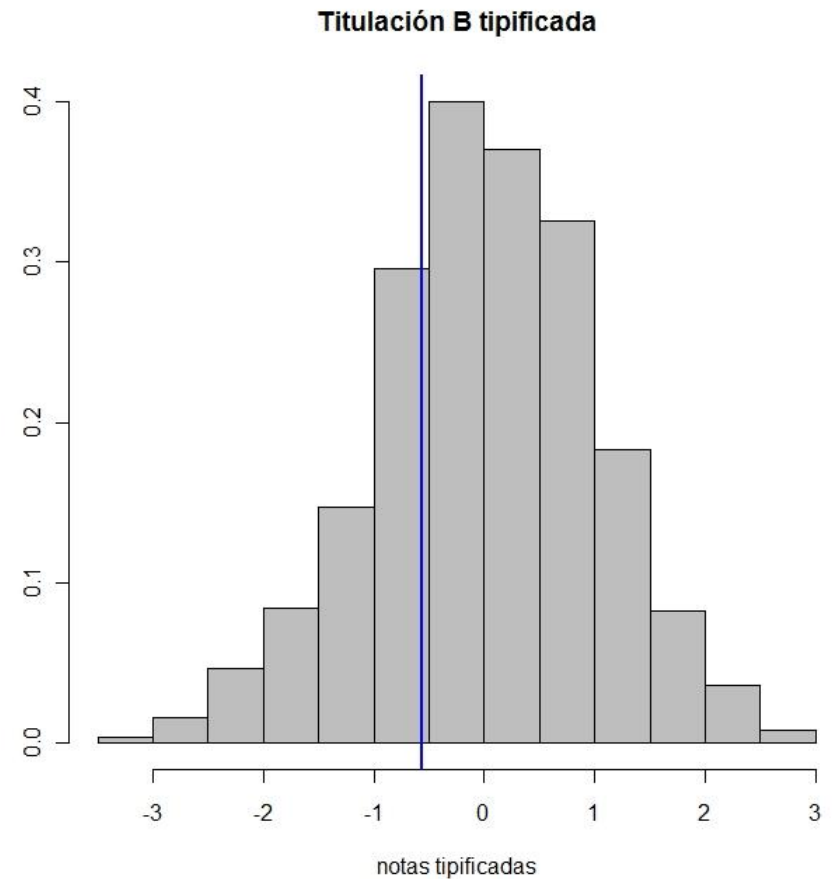
$$\frac{7.5 - 7.92}{0.75} = -0.56$$

# Medidas características

□ Alumno titulación A:



□ Alumno titulación B:



# Medidas características

- Coeficiente de asimetría de Fisher:

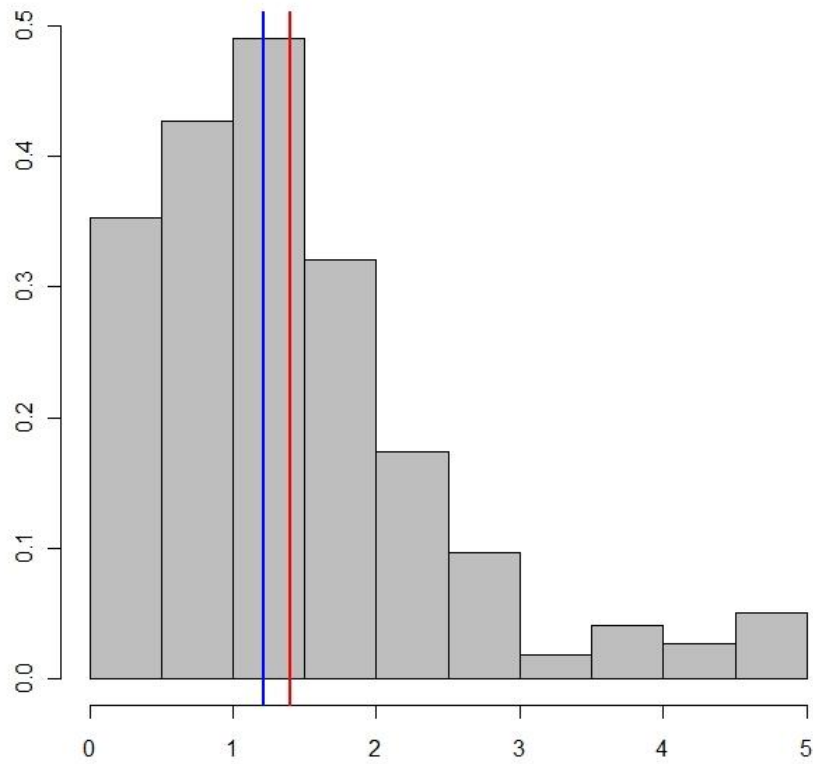
$$\gamma_1 = \frac{1}{s^3} \frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})^3$$

- Un ejemplo: curva de ingresos...
- Compárese con la Normal (simétrica, coeficiente 0)

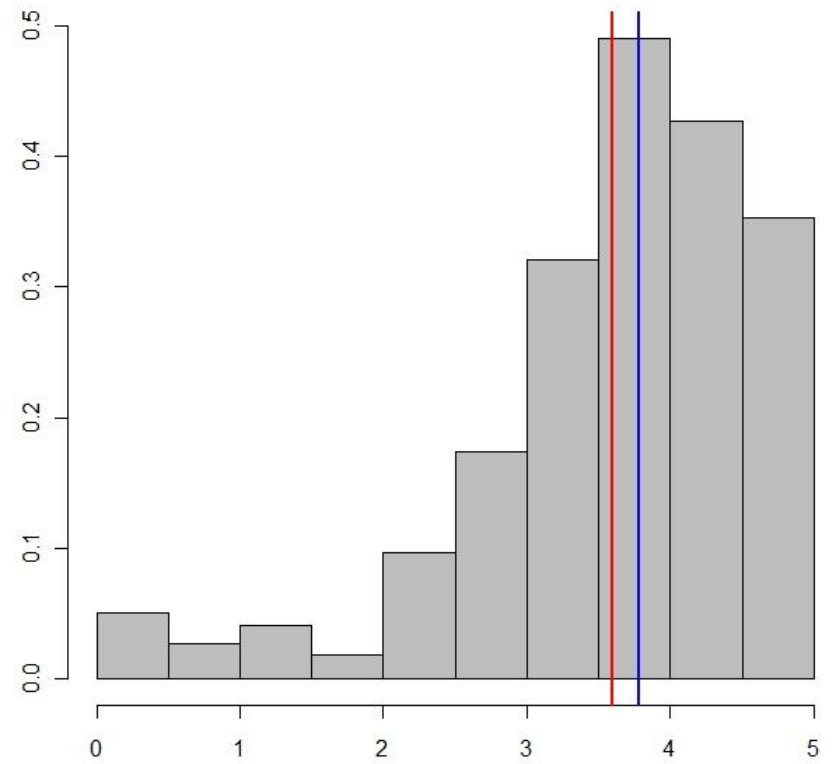
# Medidas características

## □ Asimetría:

Asimetría positiva



Asimetría negativa



# Medidas características

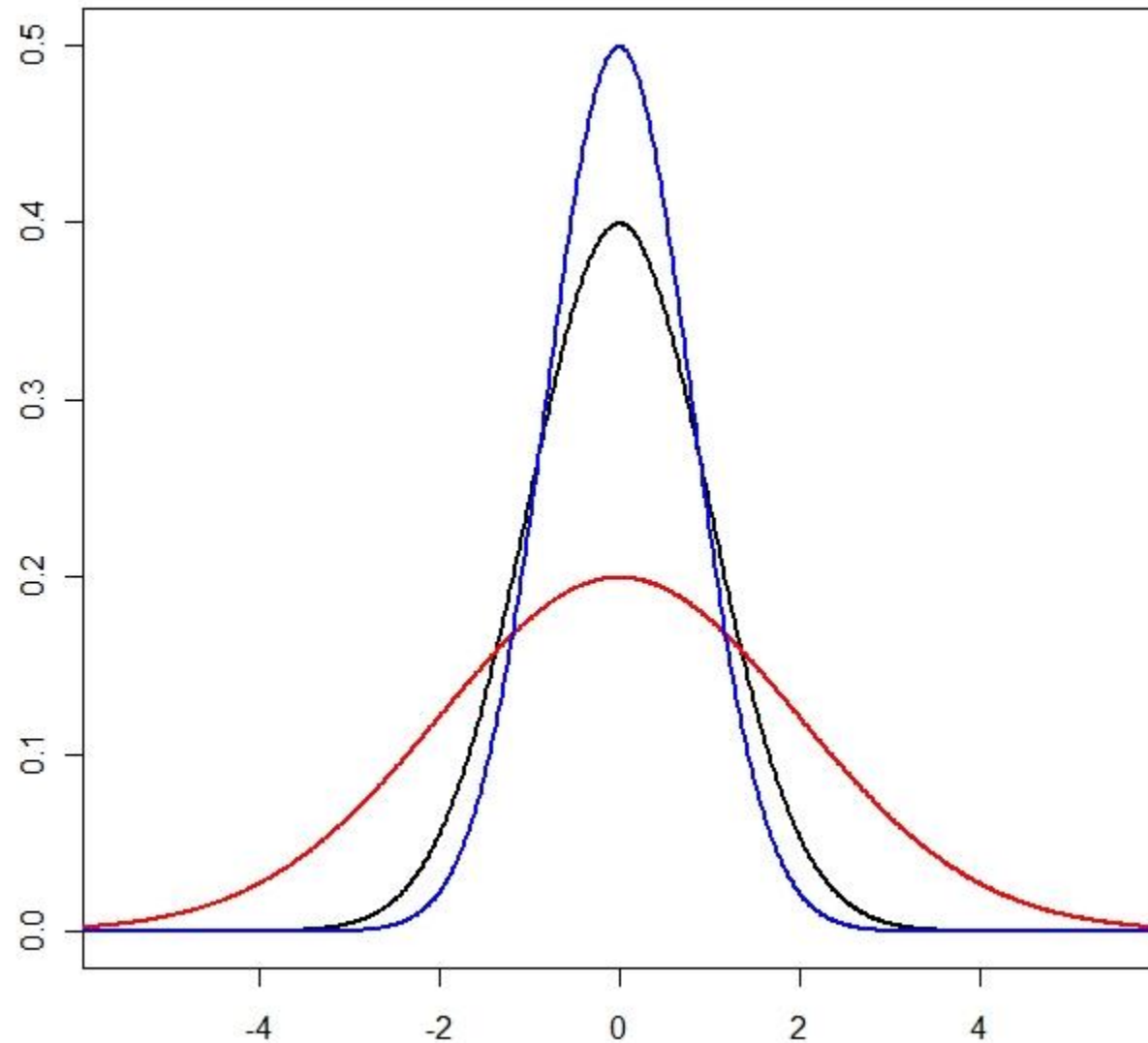
- Coeficiente de curtosis de Fisher (sólo distribuciones simétricas):

$$\gamma_2 = \frac{1}{s^4} \frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})^4 - 3$$

- Compárese con la distribución Normal o campana de Gauss (que tiene coeficiente de curtosis 0)

# Medidas características

- Curtosis:
- Leptocúrtica
- Mesocúrtica
- Planicúrtica



# Medidas características

- Momentos con respecto al origen

$$a_r = \frac{1}{n} \sum_{i=1}^n x_i^r$$

- Momentos centrales

$$m_r = \frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})^r$$

# Medidas características

## □ En particular

### □ Media:

$$\bar{x} = a_1$$

### □ Varianza:

$$s^2 = m_2$$

### □ Coef. Asimetría:

$$\gamma_1 = \frac{m_3}{s^3}$$

### □ Coef. Curtosis:

$$\gamma_2 = \frac{m_4}{s^4} - 3$$

# Módulo 1. Estadística descriptiva



Salvador Naya Fernández  
salvador.naya@udc.es

# MÓDULO 2. ESTADÍSTICA DESCRIPTIVA DE DOS VARIABLES

Salvador Naya Fernández. Universidad de A Coruña

# Programa de Probabilidades y estadística en Secundaria (España)

- **Tema 1: Estadística descriptiva unidimensional.**  
Tipos de variables. Tablas y gráficos. Parámetros estadísticos de centralización, de dispersión y de posición.
- **Tema 2: Distribuciones bidimensionales.**  
**Interpretación de partir de la representación gráfica de una nube de puntos. Grado de relación entre las variables estadísticas. Regresión lineal. Extrapolación de resultados.**
- **Tema 3: Azar y probabilidad.** Distribuciones de probabilidad binomial y normal.

# Programa de Probabilidades y estadística en Secundaria (España)

## CRITERIOS DE EVALUACIÓN

- **Utilizar las tablas y gráficas** como instrumento **para el estudio de situaciones empíricas** relacionadas con fenómenos sociales.
- **Interpretar la posible relación** entre las variables de una distribución bidimensional utilizando la recta de regresión y el coeficiente de correlación.
- **Utilizar técnicas estadísticas** elementales para **asignar probabilidades** en situaciones que se ajusten a una distribución de probabilidad binomial o normal.

# Descriptiva de dos variables

- Vector estadístico
- Tablas de frecuencias conjuntas
- Tablas de contingencia
- Representación de cualitativa-cuantitativa
- Diagrama de dispersión (cuant-cuant)
- Regresión y correlación

# descriptiva de dos variables

- Tablas de frecuencias conjuntas:
  - ▣ dos variables X e Y, como modalidades:

- Para X:

$$c_i, \quad i = 1, \dots, k$$

- Para Y:

$$d_j, \quad j = 1, \dots, l$$

# descriptiva de dos variables

- Tabla de doble entrada para las frecuencias absolutas...

$X \setminus Y$	$d_1$	$d_2$	...	$d_l$	
$c_1$	$n_{11}$	$n_{12}$	...	$n_{1l}$	$n_{1\bullet}$
$c_2$	$n_{21}$	$n_{22}$	...	$n_{2l}$	$n_{2\bullet}$
...	...	...	...	...	...
$c_k$	$n_{k1}$	$n_{k2}$	...	$n_{kl}$	$n_{k\bullet}$
	$n_{\bullet 1}$	$n_{\bullet 2}$	...	$n_{\bullet l}$	$N$

# descriptiva de dos variables

- Para las frecuencias absolutas...

$X \setminus Y$	$d_1$	$d_2$	...	$d_l$	
$c_1$	$n_{11}$	$n_{12}$	...	$n_{1l}$	$n_{1\bullet}$
$c_2$	$n_{21}$	$n_{22}$	...	$n_{2l}$	$n_{2\bullet}$
...	...	...	...	...	...
$c_k$	$n_{k1}$	$n_{k2}$	...	$n_{kl}$	$n_{k\bullet}$
	$n_{\bullet 1}$	$n_{\bullet 2}$	...	$n_{\bullet l}$	$N$



# Tablas de contingencia

	<b>Rubio</b>	<b>Pelirrojo</b>	<b>Castaño</b>	<b>Oscuro</b>	<b>Negro</b>	
<b>Claros</b>	688	116	584	188	4	<b>1580</b>
<b>Azules</b>	326	38	241	110	3	<b>718</b>
<b>Castaños</b>	343	84	909	412	26	<b>1774</b>
<b>Oscuros</b>	98	48	403	681	81	<b>1311</b>
	<b>1455</b>	<b>286</b>	<b>2137</b>	<b>1391</b>	<b>114</b>	<b>5383</b>

Datos de 5383 escolares escoceses color del pelo frente a color de ojos (Fisher, 1940)

# Tablas de contingencia

	<b>Rubio</b>	<b>Rojo</b>	<b>Castaño</b>	<b>Oscuro</b>	<b>Negro</b>	
<b>Claros</b>	688	116	584	188	4	<b>1580</b>
<b>Azules</b>	326	38	241	110	3	<b>718</b>
<b>Castaños</b>	343	84	909	412	26	<b>1774</b>
<b>Oscuros</b>	98	48	403	681	81	<b>1311</b>
	<b>1455</b>	<b>286</b>	<b>2137</b>	<b>1391</b>	<b>114</b>	<b>5383</b>

Distribuciones marginales

# Tablas de contingencia

	<b>Rubio</b>	<b>Rojo</b>	<b>Castaño</b>	<b>Oscuro</b>	<b>Negro</b>	
<b>Claros</b>	688	116	584	188	4	<b>1580</b>
<b>Azules</b>	326	38	241	110	3	<b>718</b>
<b>Castaños</b>	343	84	909	412	26	<b>1774</b>
<b>Oscuros</b>	98	48	403	681	81	<b>1311</b>
	<b>1455</b>	<b>286</b>	<b>2137</b>	<b>1391</b>	<b>114</b>	<b>5383</b>

Distribuciones marginales

# Tablas de contingencia

- Resultados de un examen de Estadística:

	Suspenseo	Aprobado	
Mujer	14	36	<b>50</b>
Hombre	20	11	<b>31</b>
	<b>34</b>	<b>47</b>	<b>81</b>

- Aprueban más las mujeres que los hombres?

# Tablas de contingencia

- Si el sexo no influye, los porcentajes de aprobados en los dos grupos deberían ser los mismos

		Suspense	Aprobado	
Mujer	Observados	14	36	50
	<b>Esperados</b>	<b>21</b>	<b>29</b>	
Hombre	Observados	20	11	31
	<b>Esperados</b>	<b>13</b>	<b>18</b>	
		34	47	<b>81</b>

- Esperaríamos,  $47 \cdot 50 / 81$  y  $47 \cdot 31 / 81$ , para hombres y mujeres, respectivamente.

# Tablas de contingencia

- una forma de medir la distancia entre lo observado y lo esperado (si las variables son independientes):

$$\chi^2 = \sum \frac{(\textit{observado} - \textit{esperado})^2}{\textit{esperado}}$$

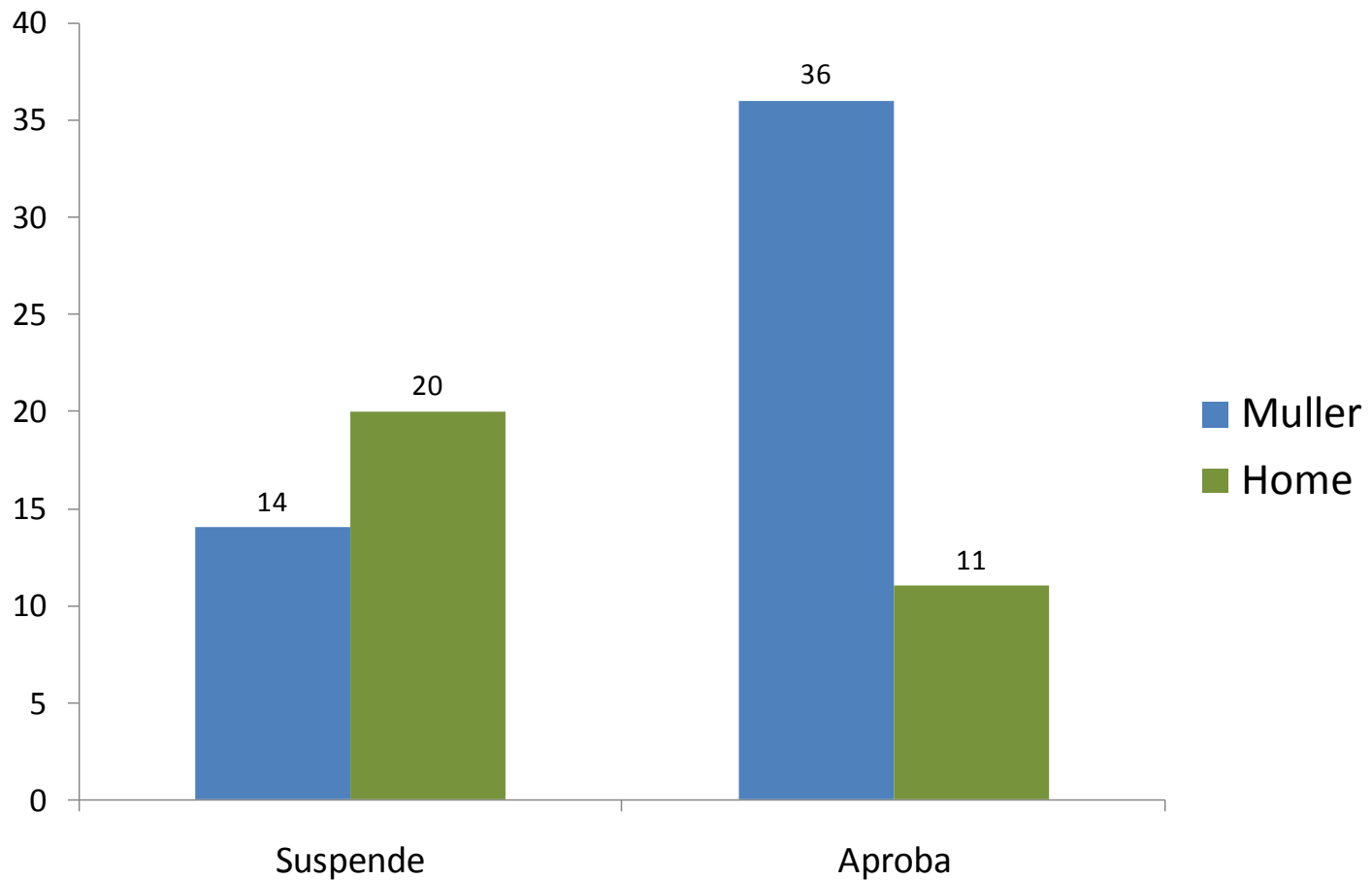
- Si el valor de este estadístico es grande, entonces hay dependencia entre las variables. En este caso: 10.51
- ... pero cuando es grande? (Inferencia: Contrastes)

# Representaciones gráficas

- Para representar cualitativa-cualitativa
  - ▣ Barras agrupadas (frecuencias absolutas)
  - ▣ Barras apiladas (frecuencias relativas)
  - ▣ Gráficos de áreas

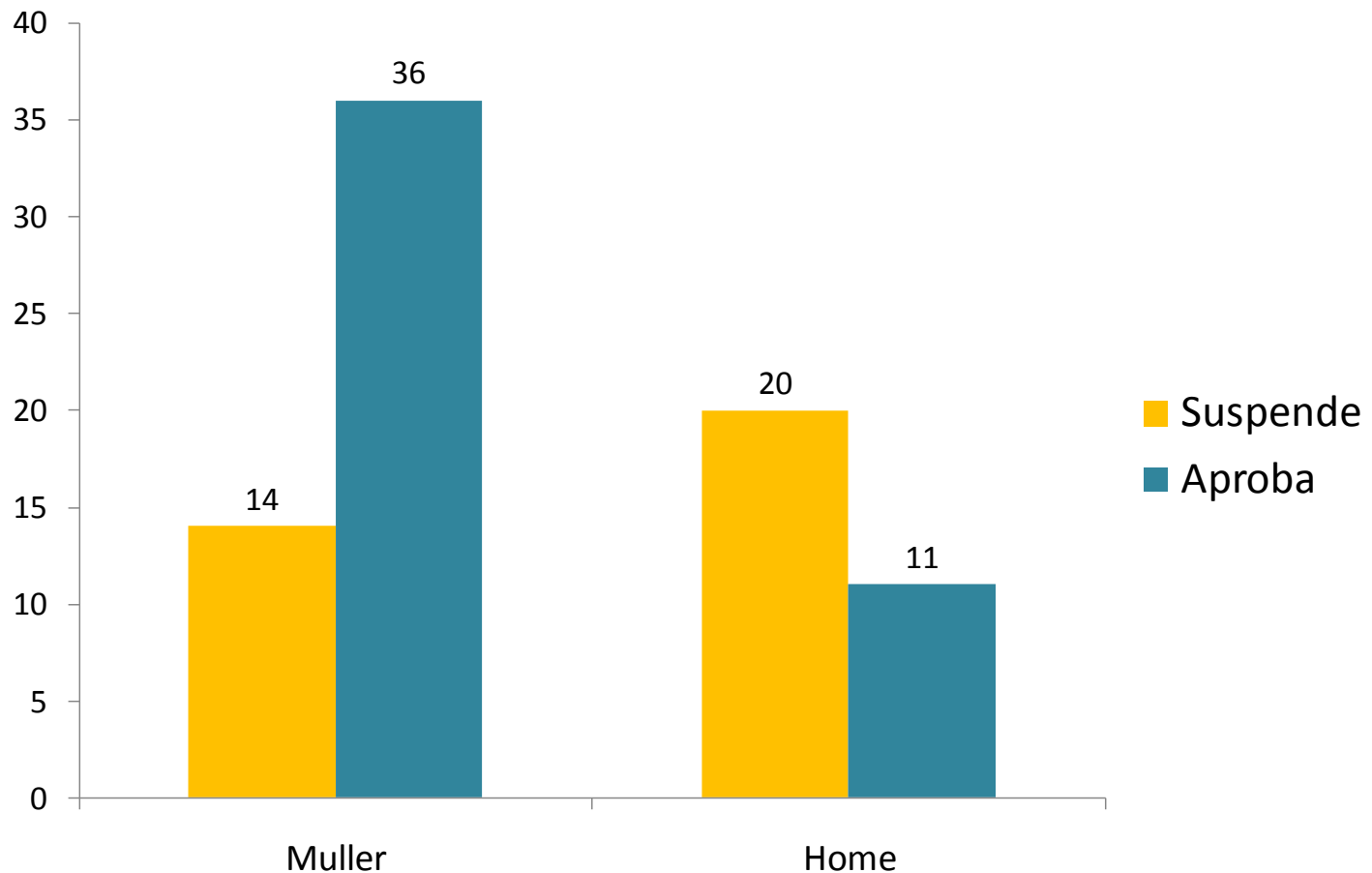
# Representaciones gráficas

## □ Barras agrupadas:



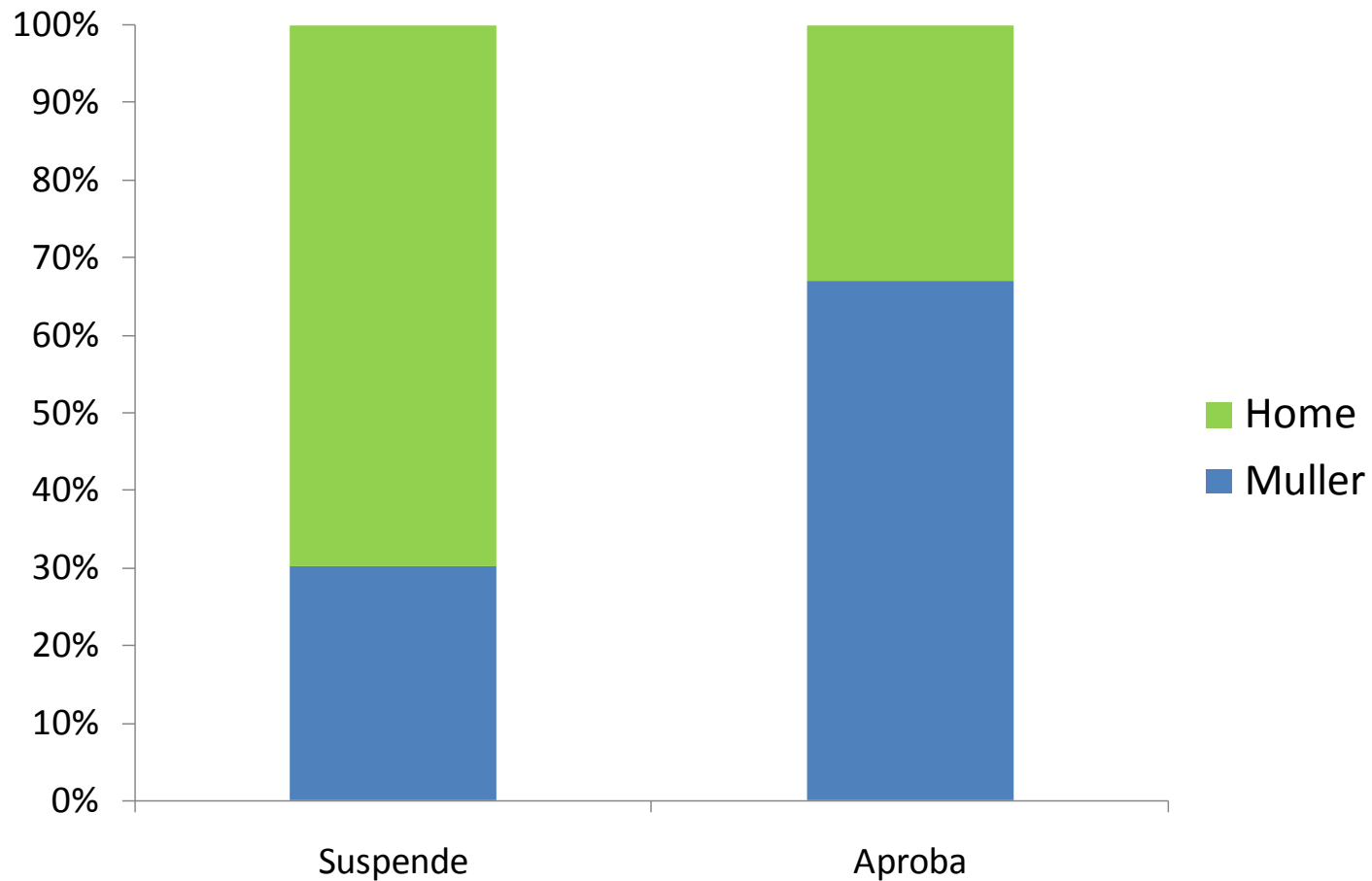
# Representaciones gráficas

## □ Barras agrupadas:



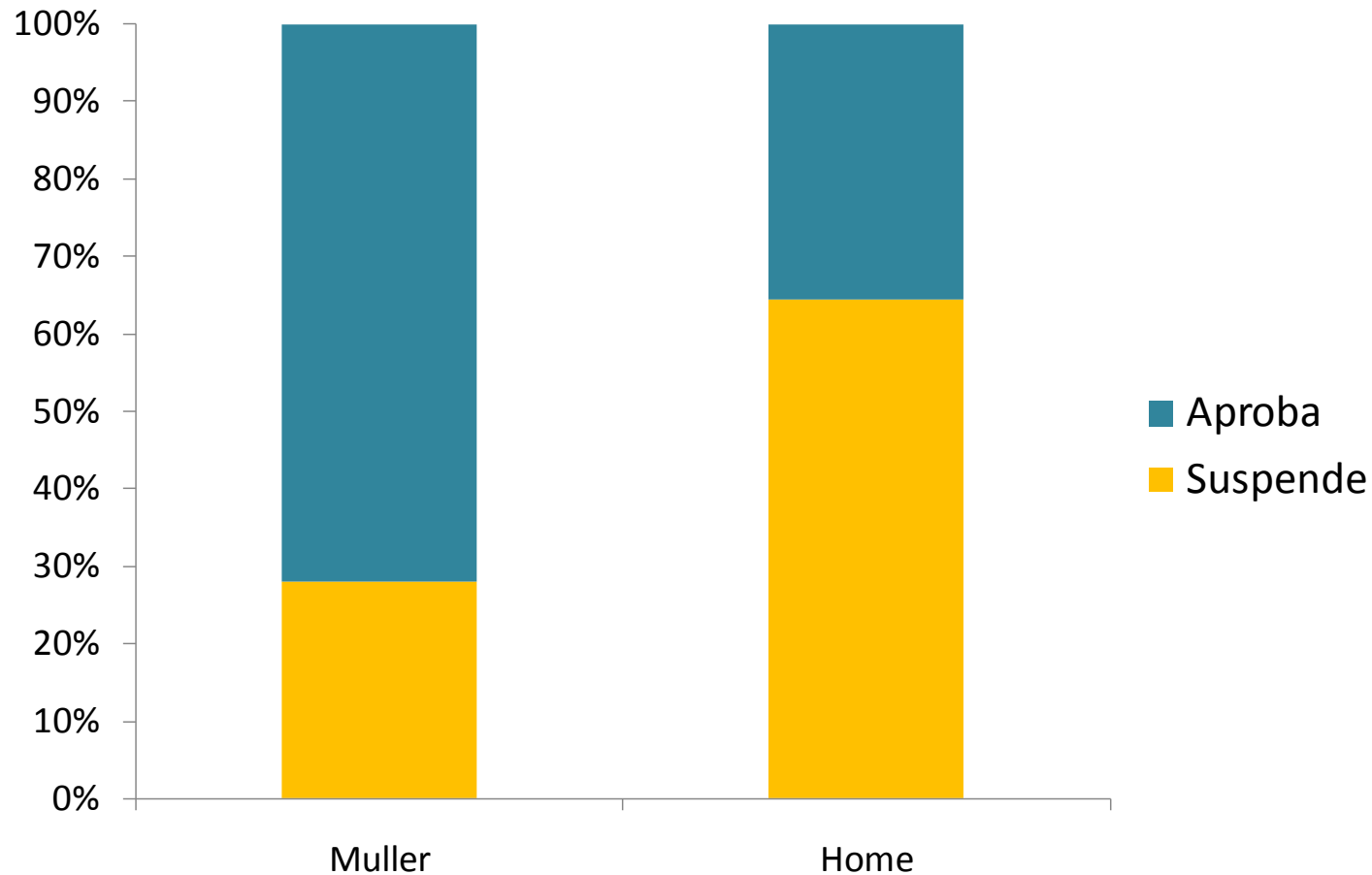
# Representaciones gráficas

## □ Barras apiladas



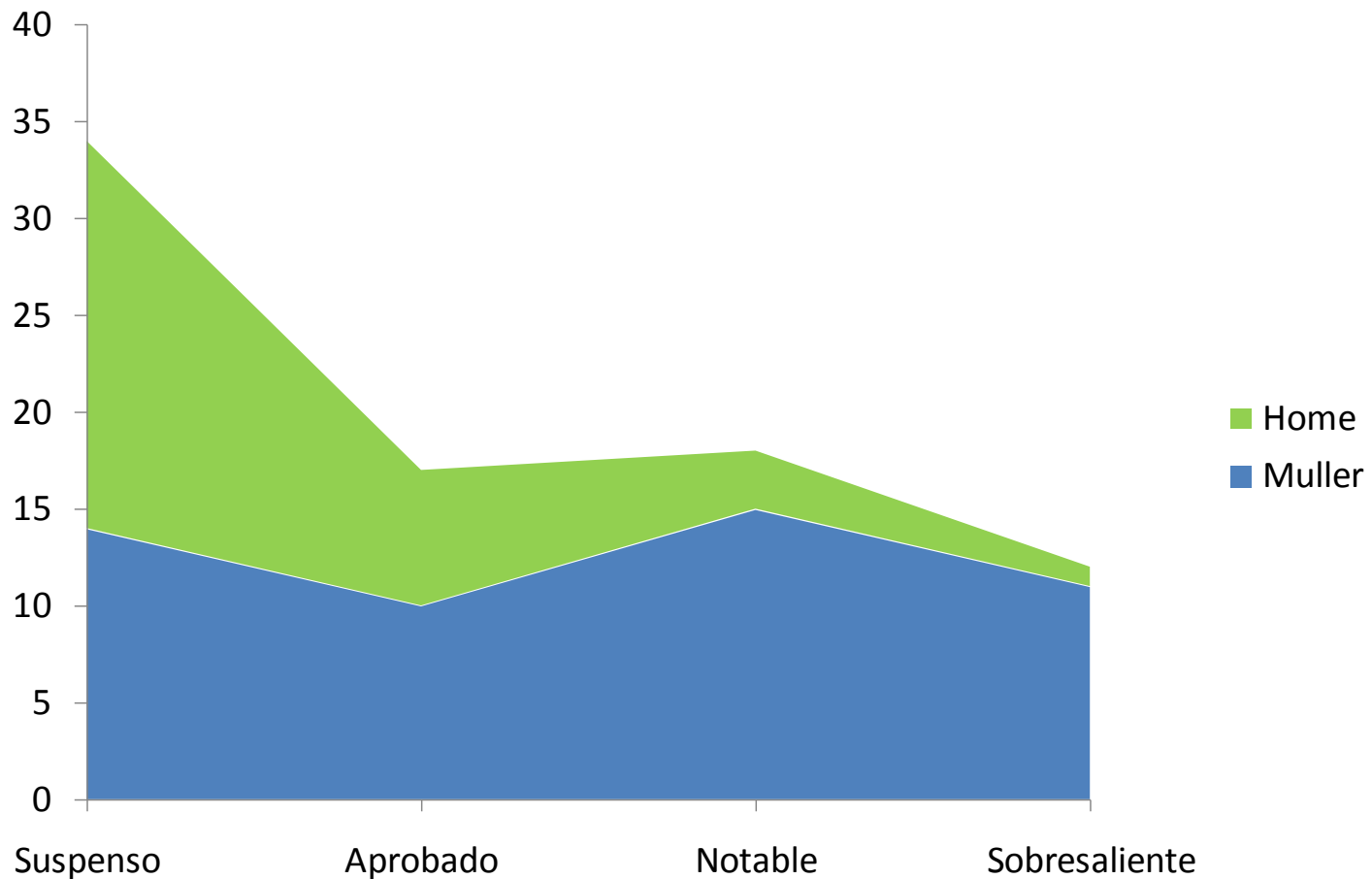
# Representaciones gráficas

## □ Barras apiladas



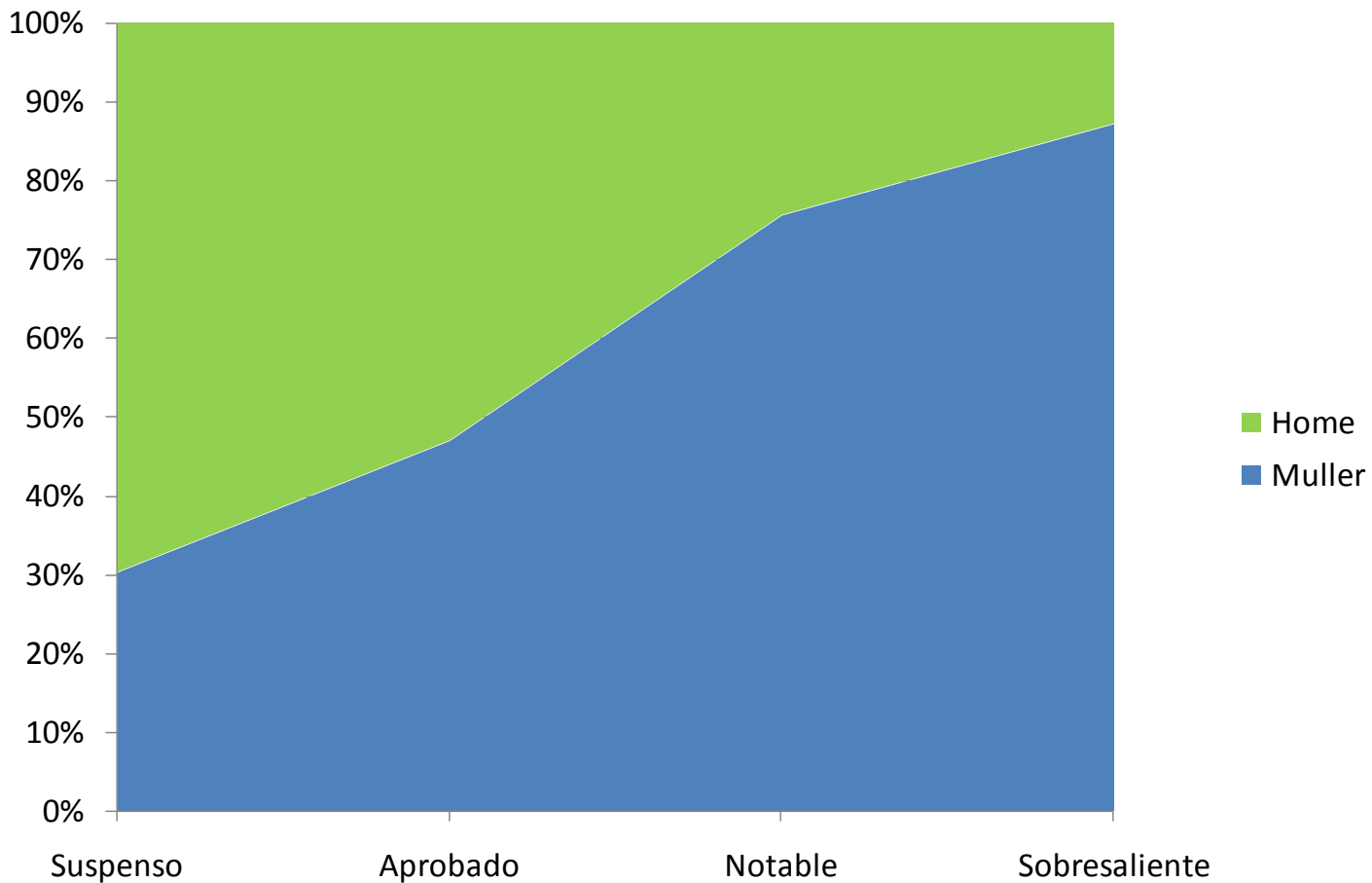
# Representaciones gráficas

## □ Gráficos de áreas



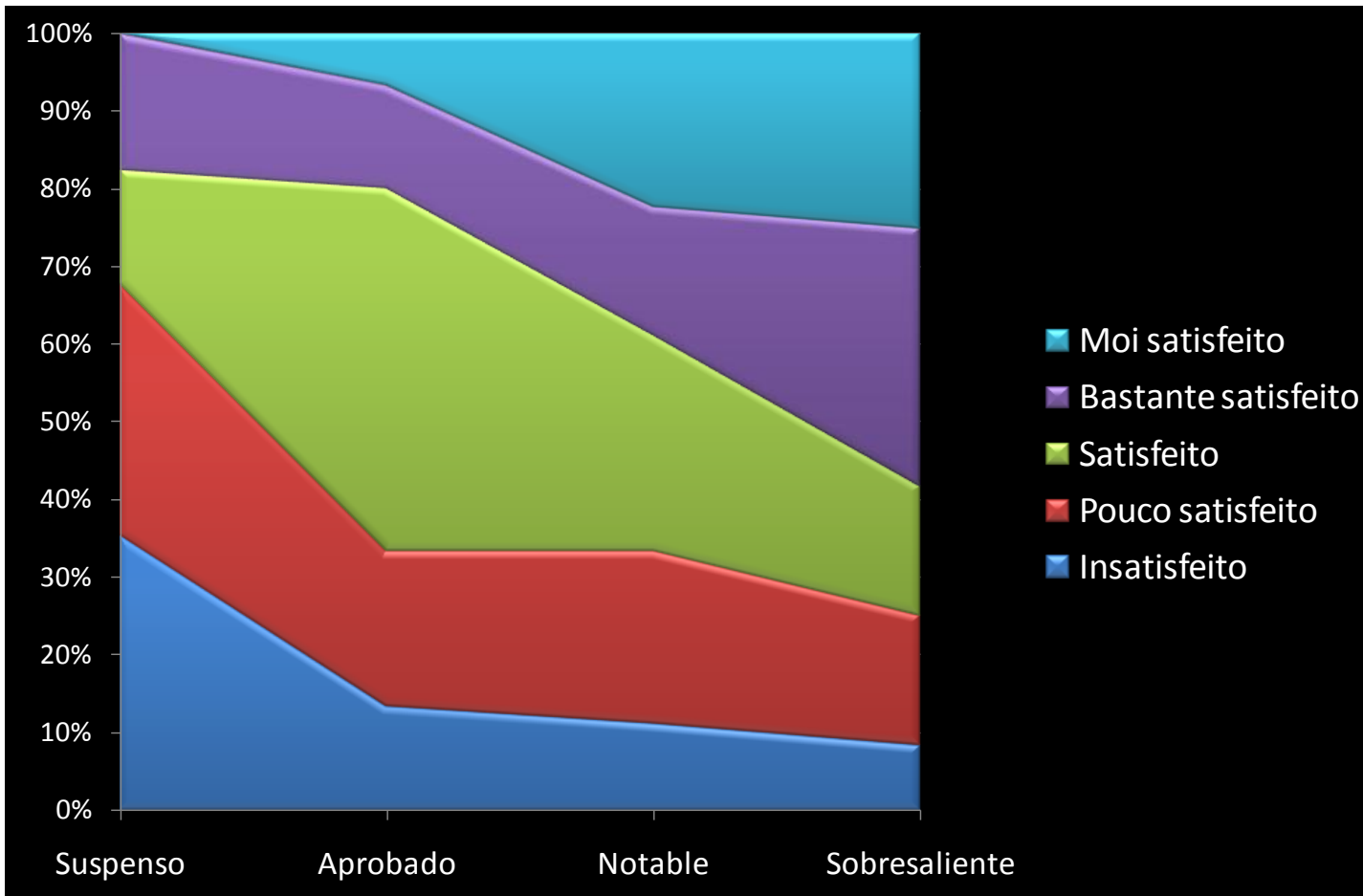
# Representaciones gráficas

## □ Gráficos de áreas



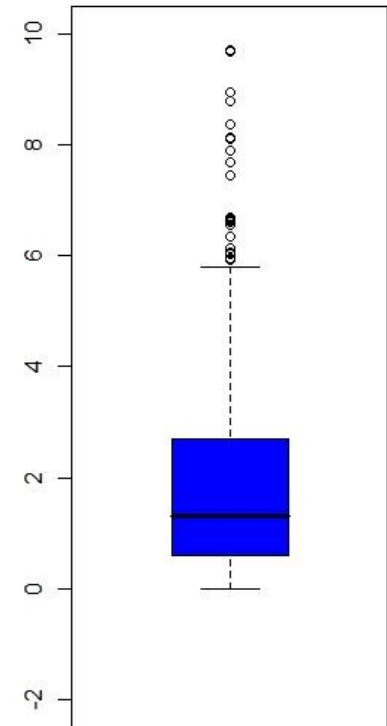
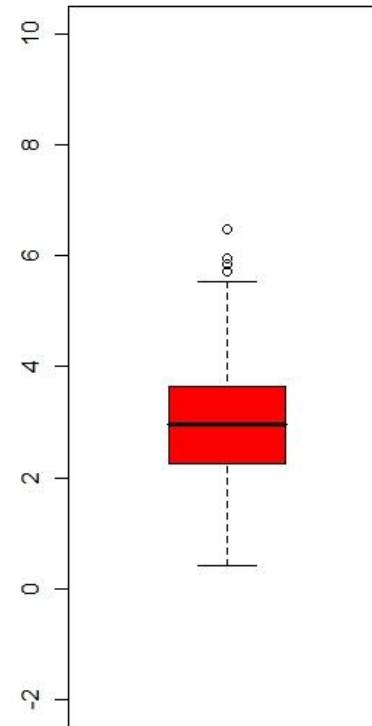
# Representaciones gráficas

## □ Gráficos de áreas



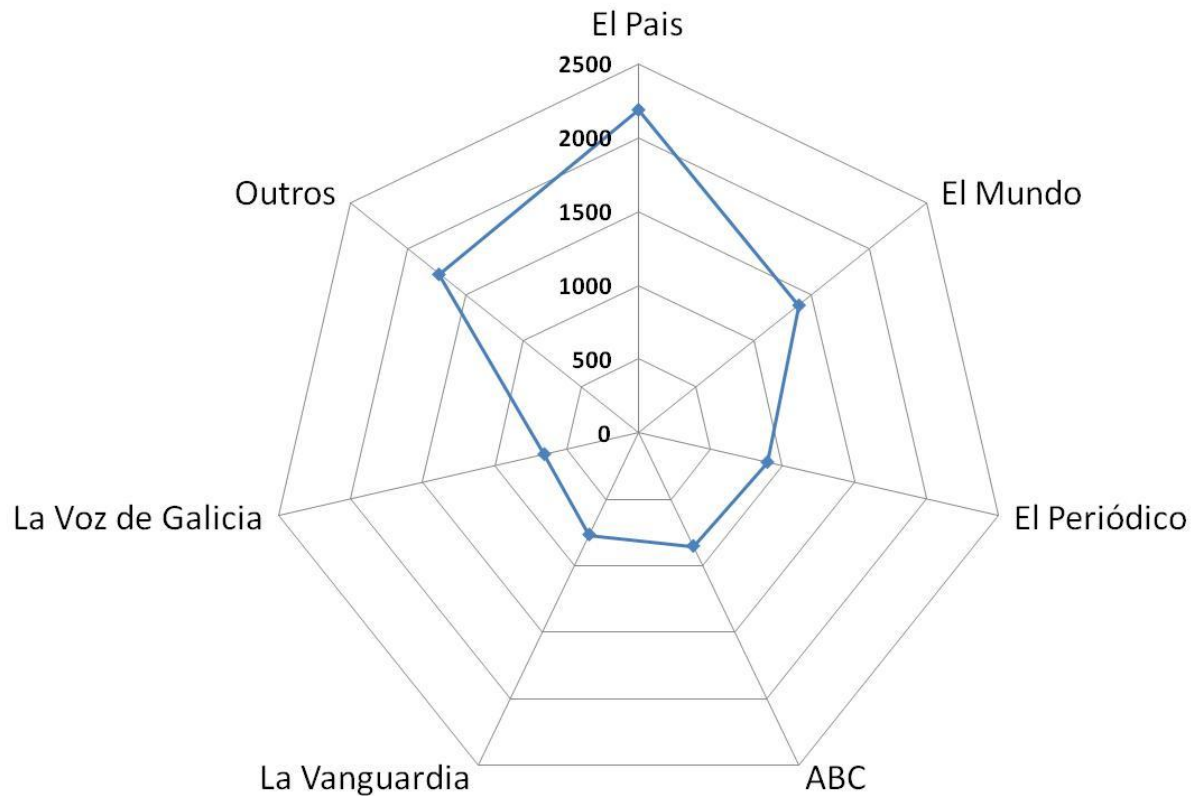
# Representaciones gráficas

- Para representar cualitativa-cuantitativa:
  - ▣ Diagramas de caja
  - ▣ Gráficos radiales



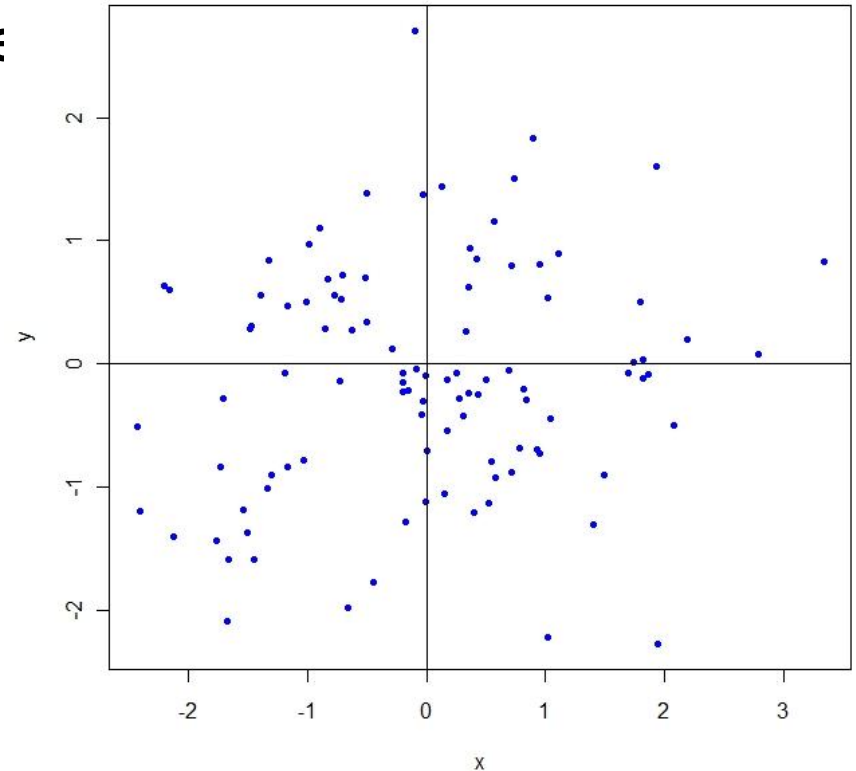
# Representaciones gráficas

## □ Gráfico radial



# Regresión y correlación

- Para *confrontar* los valores de dos variables continuas, se utiliza el diagrama de dispersión o nube de puntos.



# Regresión y correlación

- Consideremos  $(X,Y)$  un vector estadístico:
  - ▣ Vector de medias:

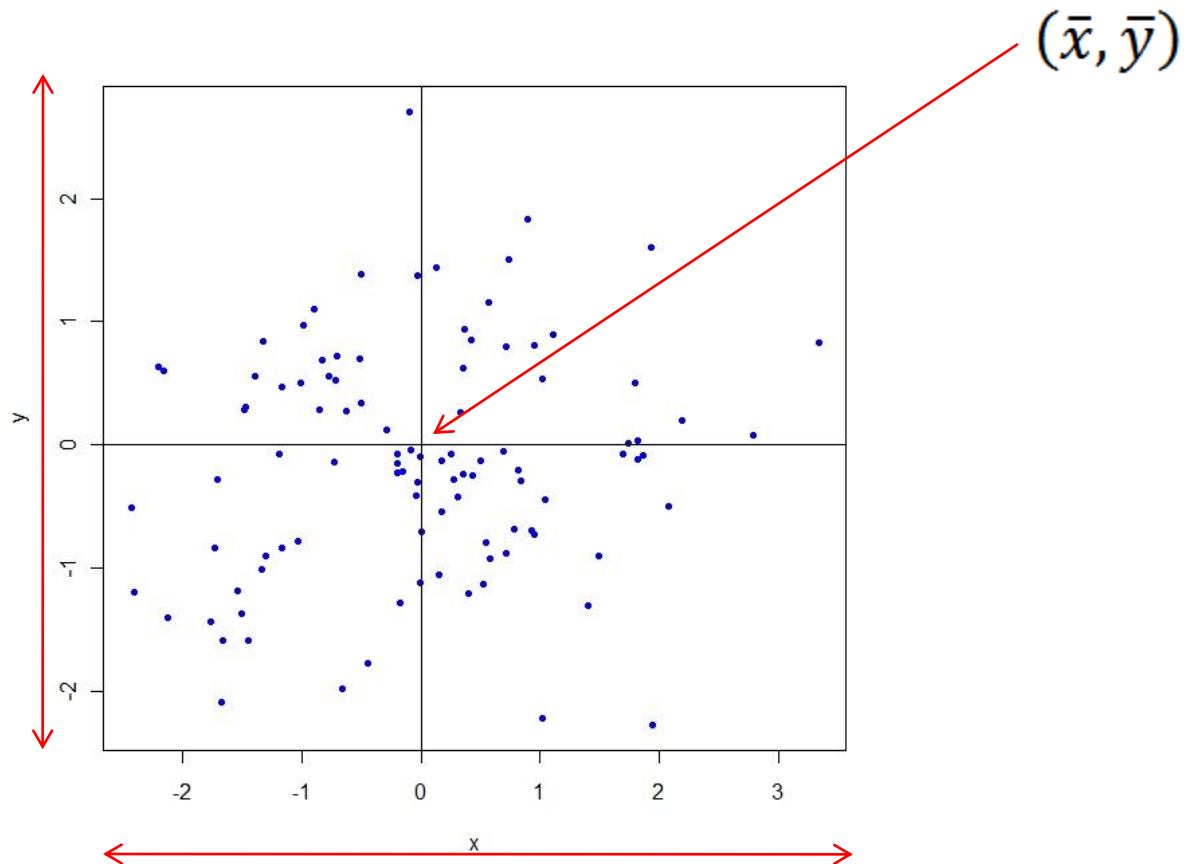
$$(\bar{x}, \bar{y})$$

- ▣ Matriz de varianzas-covarianzas:

$$\Sigma = \begin{pmatrix} S_x^2 & S_{xy} \\ S_{xy} & S_y^2 \end{pmatrix}$$

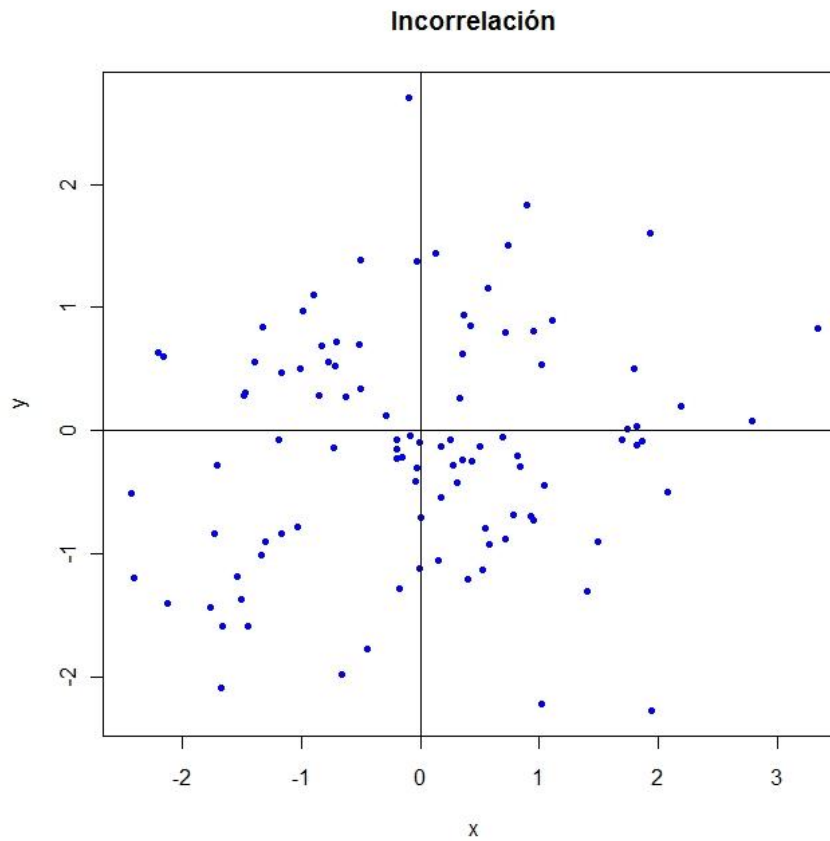
# Regresión y correlación

## □ Vector de medias



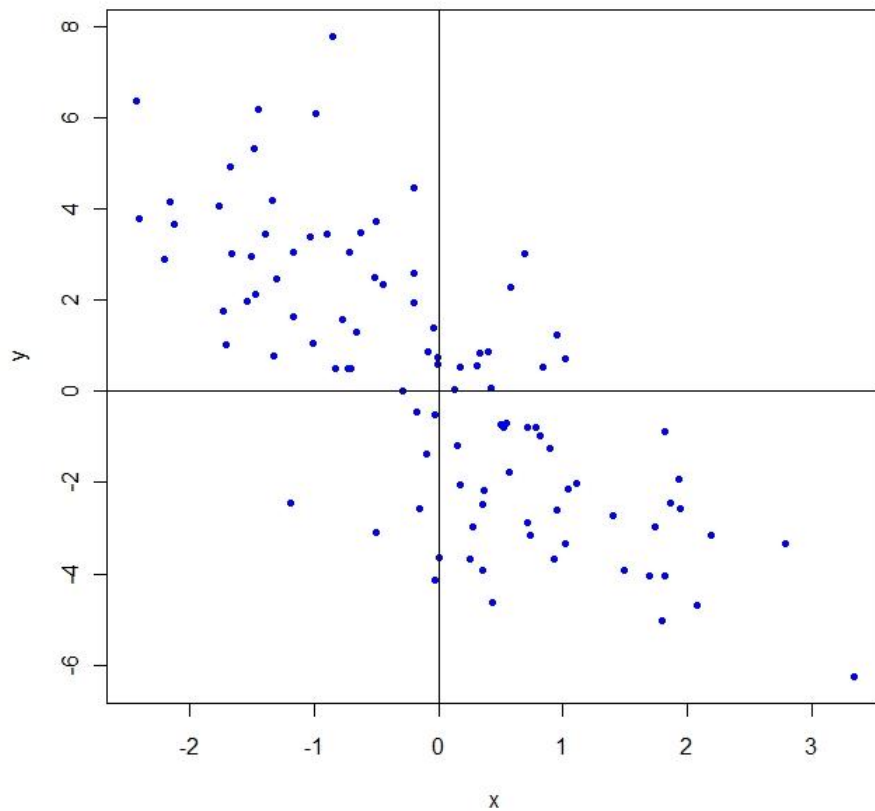
# Regresión y correlación

## □ Tipos de relación

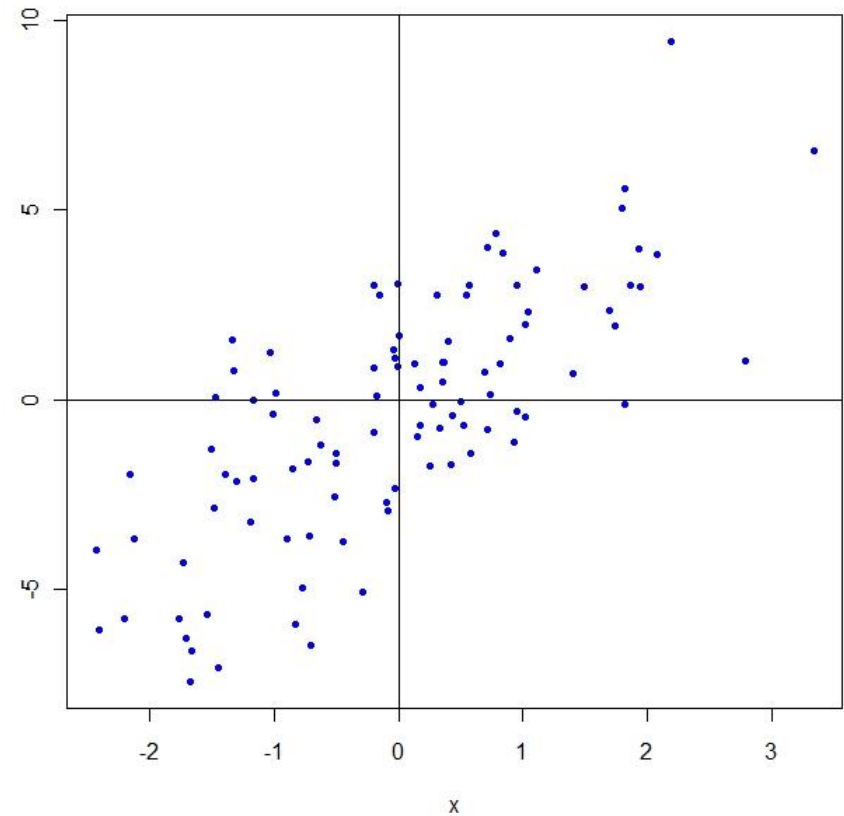


# Regresión y correlación

Relación inversa



Relación directa



# Regresión y correlación

- La **covarianza** entre dos variables indica si existe relación entre ellas, y si esta es directa o inversa.

$$S_{xy} = \frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})$$

- Que también se puede calcular como:

$$S_{xy} = \frac{1}{n} \sum_{i=1}^n x_i y_i - \bar{x}\bar{y}$$

# Regresión y correlación

- La **covarianza** entre dos variables indica se existe relación entre ellas, e se esta é directa o inversa.

$$S_{xy} = \frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})$$

- El coeficiente de correlación lineal de Pearson de dos variables,  $r$ , indica si los puntos tienden a disponerse alineadamente:

$$r = \frac{S_{xy}}{S_x S_y}$$



# Regresión y correlación

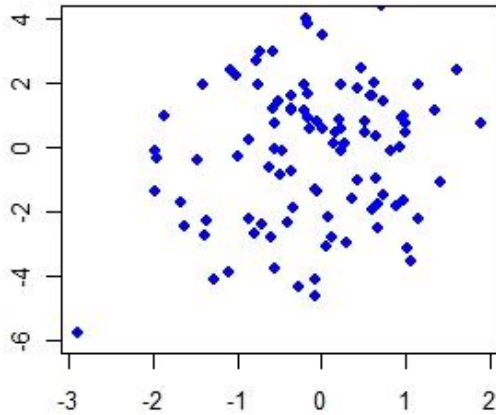
## □ Propiedades:

- Mismo signo que la covarianza
- Útil para determinar relaciones lineales
- Es adimensional
- Toma valores en  $[-1, 1]$ 
  - $r=0$  : variables incorreladas
  - $r \sim 1$ : relación lineal directa
  - $r \sim (-1)$ : relación lineal inversa

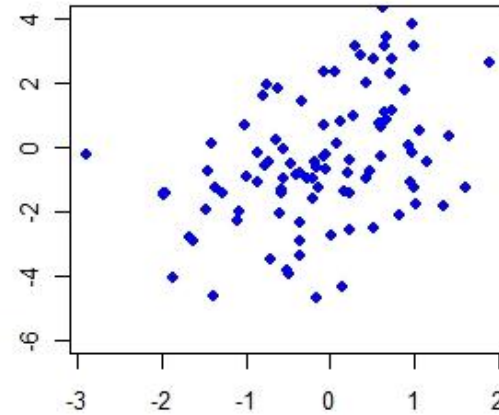
Y se  $r=0$ , son independientes?

# Regresión y correlación

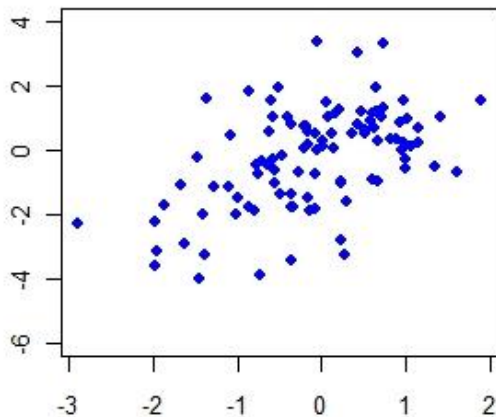
$\rho=0.12$



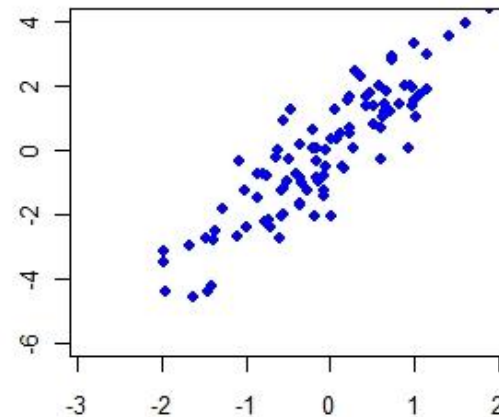
$\rho=0.35$



$\rho=0.51$



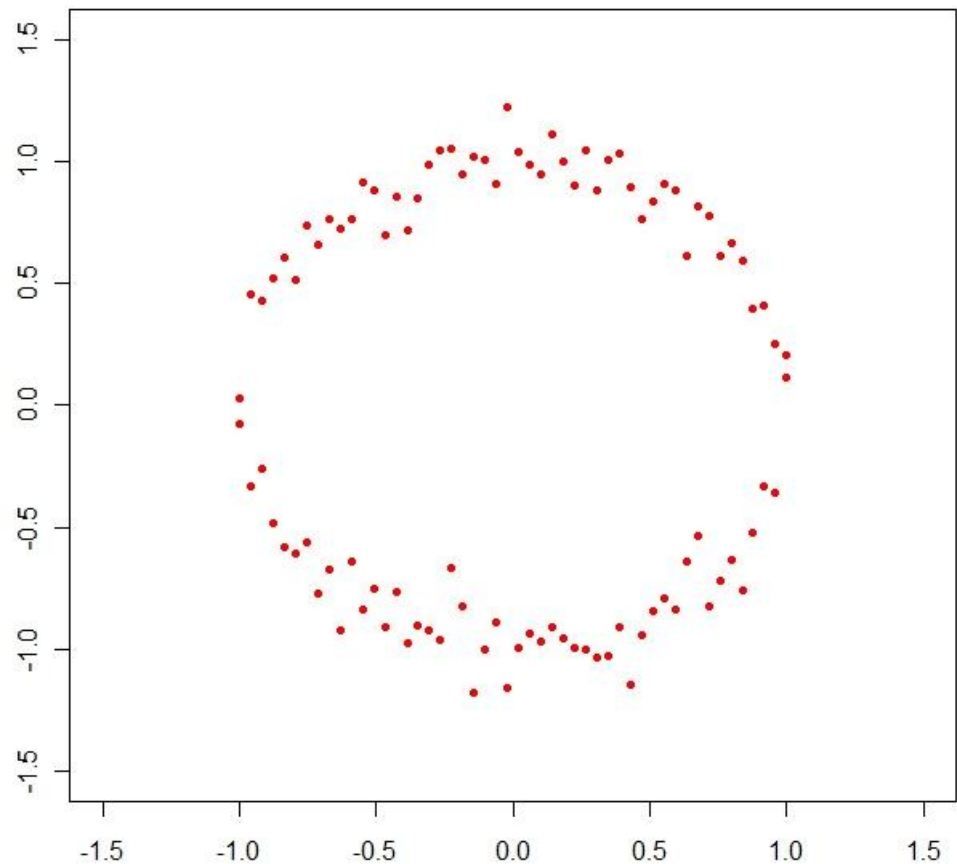
$\rho=0.91$



# Regresión y correlación

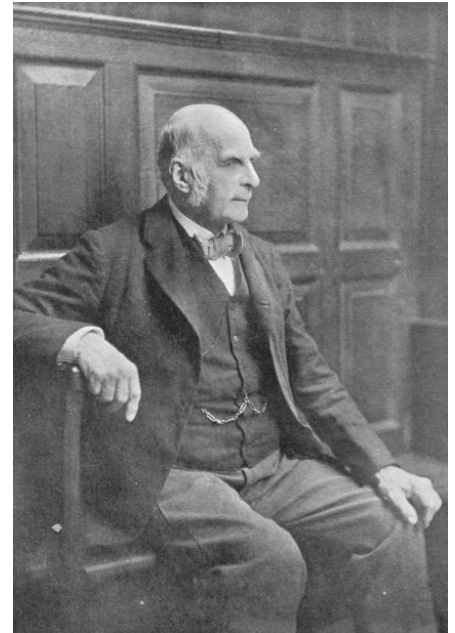
- Si son incorrelados...  
son independientes?  
No necesariamente:

Incorrelados pero dependientes



# Regresión y correlación

- El término *regresión* fue introducido por Galton (padre de la BioEstadística) en su publicación *Natural inheritance* (1889), en referencia a la ley de regresión natural.
- “Cada peculiaridad en un hombre es compartida por sus descendientes, pero en media, en un grado menor.”
  - Regresión a la media



# Regresión y correlación

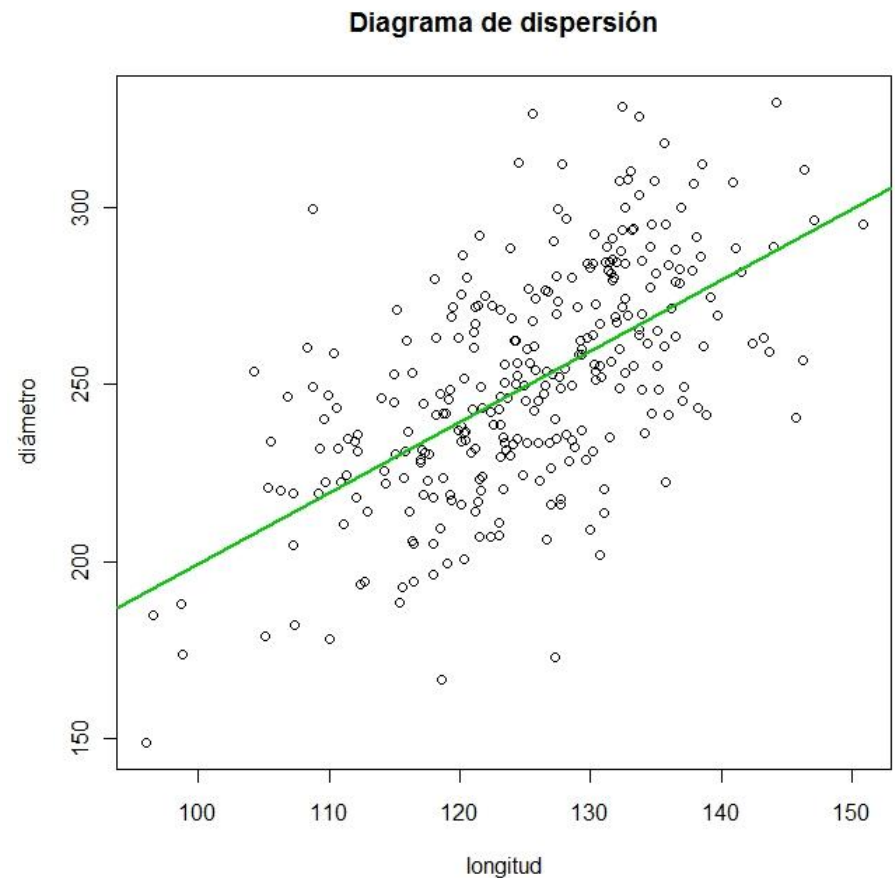


- una fábrica produce este tipo de tornillos de los que tomamos una muestra para saber cuál es la relación entre la longitud y el diámetro.

# Regresión y correlación

- Diagrama de dispersión para a longitud ( $X$ ) frente al diámetro ( $Y$ ).
- Identificaremos:
  - ▣  $X$ : variable independiente
  - ▣  $Y$ : variable dependiente
  - ▣ Buscamos:

$$Y = m(X) + \varepsilon$$



# Regresión y correlación

- Modelo “inferencial”:

$$Y = m(X) + \varepsilon$$

$$m(X) = a + bX$$

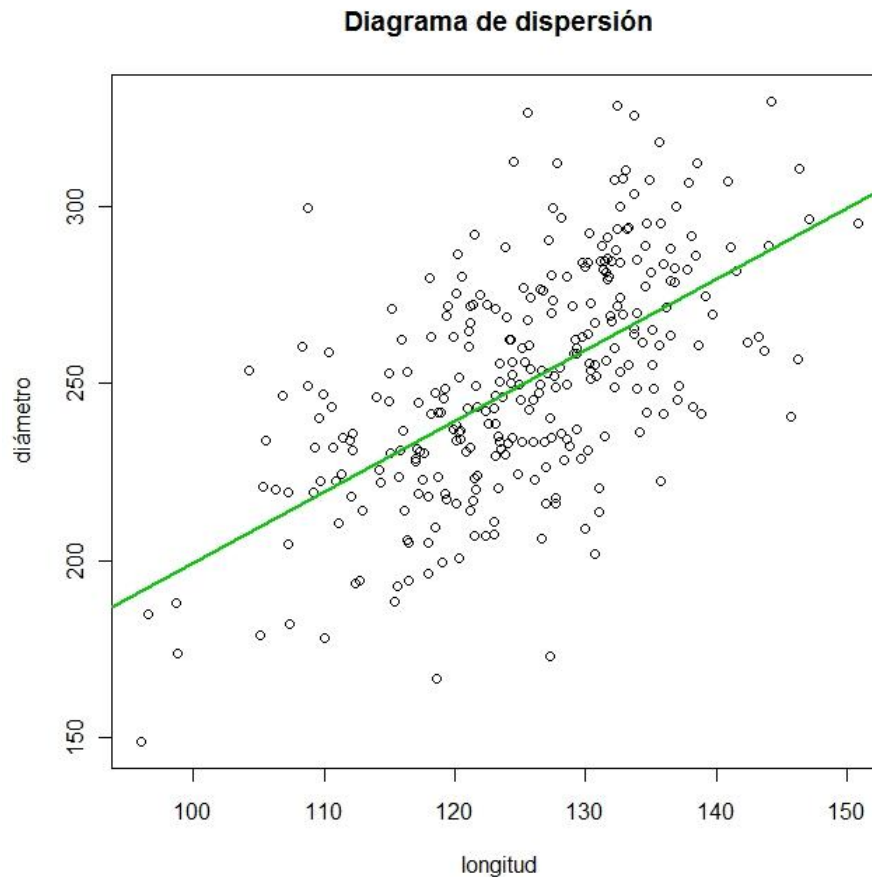
- Modelo descriptivo: no tenemos en cuenta las Distribuciones de las variables que intervienen en el modelo; sólo las observaciones muestrales.

$$y = m(x) + e$$

$$m(x) = a + bx$$

# Regresión y correlación

- Cómo ajustamos la recta a la nube de puntos?

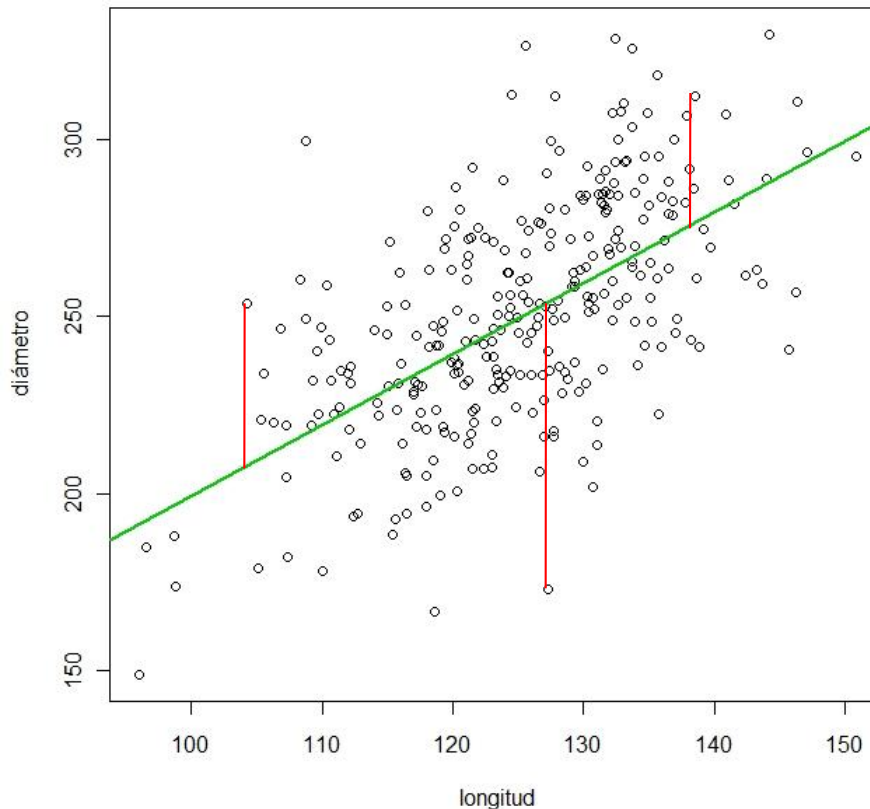


$$y = a + bx$$

# Regresión y correlación

- Cómo ajustamos la recta a la nube de puntos?

Diagrama de dispersión



$$y = a + bx$$

residuos

Objetivo: hacer que las **diferencias** entre los valores observados y los que proporciona el modelo sean *pequeñas*.

# Regresión y correlación

- Método de mínimos cuadrados (Gauss):
  - ▣ Ajustamos una recta (también válido para cualquier función paramétrica).
  - ▣ La recta ajustada dará una *aproximación* de cada valor de la variable  $Y$ .
  - ▣ Objetivo: minimizar los errores (al cuadrado) de la *aproximación*.

$$\left(\hat{a}, \hat{b}\right) = \arg \min_{(a,b)} \sum_{i=1}^n \left(y_i - a - bx_i\right)^2$$

# Regresión y correlación

## □ Método de mínimos cuadrados:

### □ El resultado:

$$b = r \frac{S_y}{S_x} = \frac{S_{xy}}{S_x^2} \quad a = \bar{y} - b\bar{x}$$

### □ La recta ajustada es:

$$y = \bar{y} + \frac{S_{xy}}{S_x^2} (x - \bar{x})$$

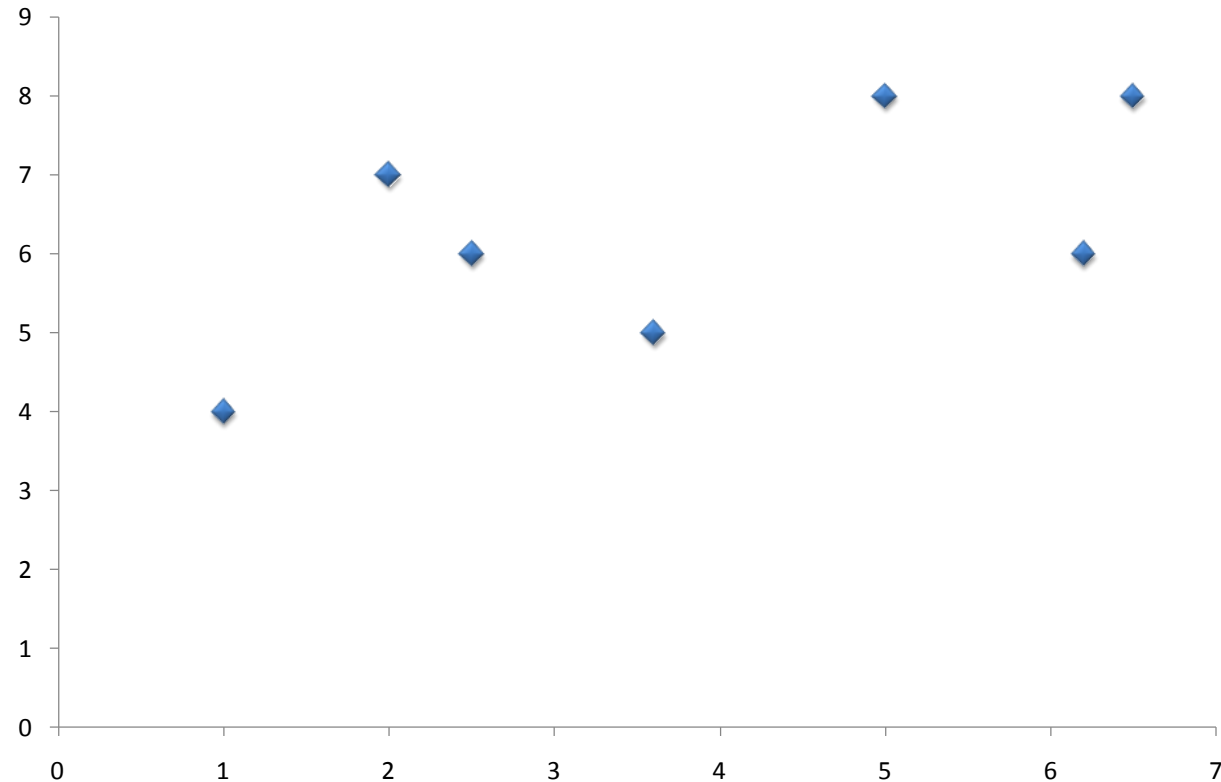
### □ $b$ es el llamado coeficiente de regresión

- Tiene el mismo signo que la covarianza

# Regresión y correlación

□ Un ejemplo: (N=7 datos)

X	Y
1	4
2	7
2,5	6
3,6	5
5	8
6,2	6
6,5	8



# Regresión y correlación

□ Un ejemplo:

X	Y	X <sup>2</sup>	Y <sup>2</sup>	XY
1	4	1	16	4
2	7	4	49	14
2,5	6	6,25	36	15
3,6	5	12,96	25	18
5	8	25	64	40
6,2	6	38,44	36	37,2
6,5	8	42,25	64	52
26,8	44	129,9	290	180,2

# Regresión y correlación

□ Un ejemplo:

X	Y	X <sup>2</sup>	Y <sup>2</sup>	XY
1	4	1	16	4
2	7	4	49	14
2,5	6	6,25	36	15
3,6	5	12,96	25	18
5	8	25	64	40
6,2	6	38,44	36	37,2
6,5	8	42,25	64	52
26,8	44	129,9	290	180,2

$$\bar{x} = \frac{26.8}{7} = 3.83$$

$$\bar{y} = \frac{44}{7} = 6.29$$

# Regresión y correlación

□ Un ejemplo:

X	Y	X <sup>2</sup>	Y <sup>2</sup>	XY
1	4	1	16	4
2	7	4	49	14
2,5	6	6,25	36	15
3,6	5	12,96	25	18
5	8	25	64	40
6,2	6	38,44	36	37,2
6,5	8	42,25	64	52
26,8	44	129,9	290	180,2

$$S_x^2 = \frac{129.9}{7} - 3.83^2 = 3.90$$

$$S_y^2 = \frac{290}{7} - 6.29^2 = 1.92$$

# Regresión y correlación

□ Un ejemplo:

X	Y	X <sup>2</sup>	Y <sup>2</sup>	XY
1	4	1	16	4
2	7	4	49	14
2,5	6	6,25	36	15
3,6	5	12,96	25	18
5	8	25	64	40
6,2	6	38,44	36	37,2
6,5	8	42,25	64	52
26,8	44	129,9	290	180,2

$$S_{xy}^2 = \frac{180.2}{7} - 3.83 \times 6.29$$
$$= 1.68$$

# Regresión y correlación

□ Un ejemplo:

X	Y	X <sup>2</sup>	Y <sup>2</sup>	XY
1	4	1	16	4
2	7	4	49	14
2,5	6	6,25	36	15
3,6	5	12,96	25	18
5	8	25	64	40
6,2	6	38,44	36	37,2
6,5	8	42,25	64	52
26,8	44	129,9	290	180,2

$$b = \frac{S_{xy}}{S_x^2} = \frac{1.68}{3.90} = 0.43$$

$$\begin{aligned} a &= \bar{y} - b\bar{x} \\ &= 6.29 - 0.43 \times 3.83 \\ &= 4.64 \end{aligned}$$

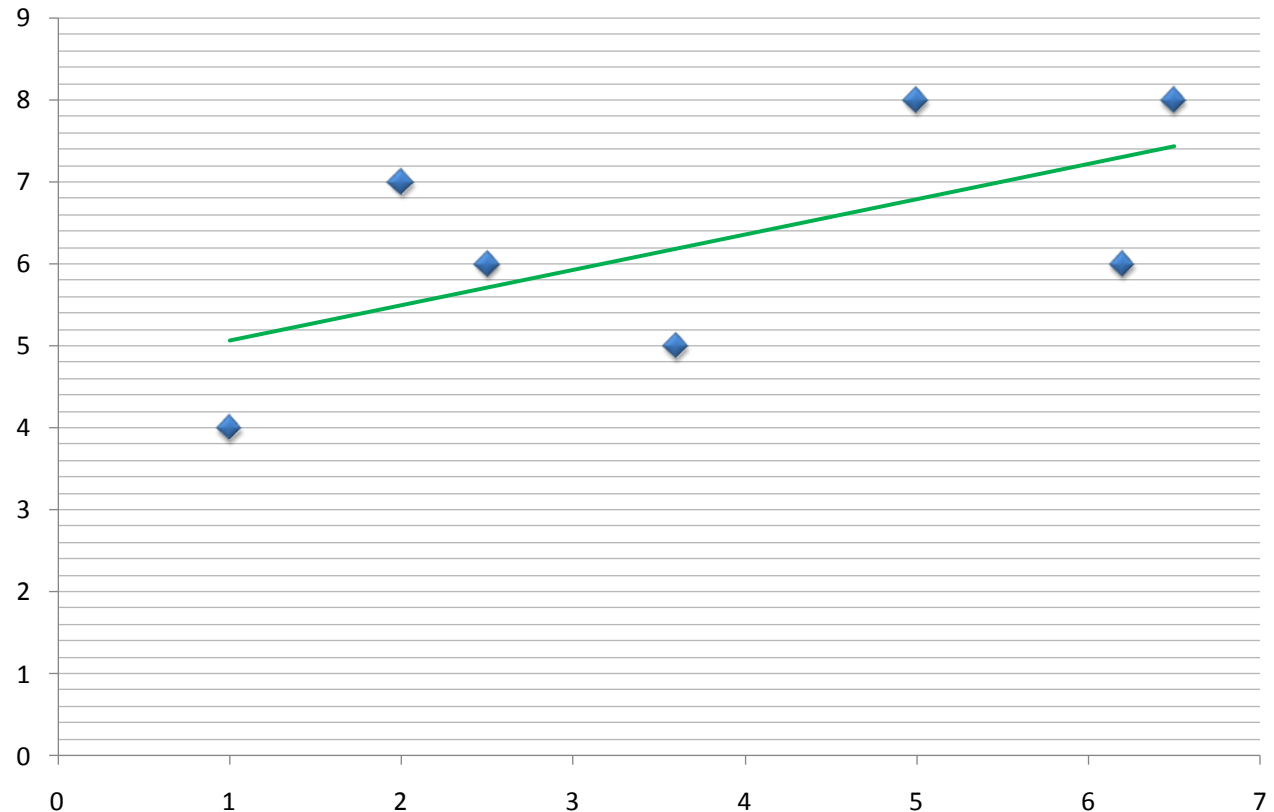
$$y = 4.64 + 0.43x$$

# Regresión y correlación

□ Un ejemplo:

$$y = 4.64 + 0.43x$$

X	Y
1	4
2	7
2,5	6
3,6	5
5	8
6,2	6
6,5	8

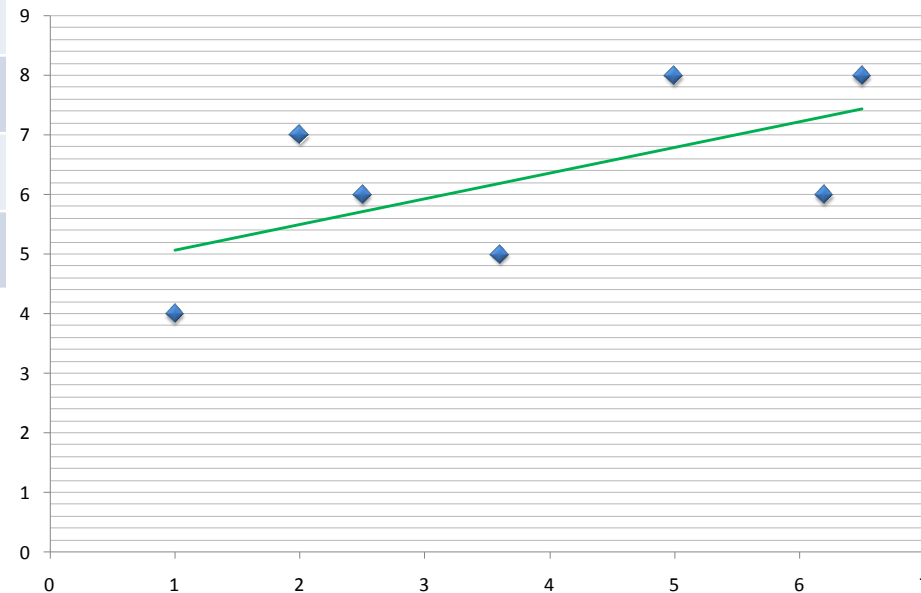


# Regresión y correlación

## □ Sobre los residuos

X	Y	Y <sup>^</sup>	e
1	4	5,07	-1,07
2	7	5,50	1,50
2,5	6	5,71	0,29
3,6	5	6,19	-1,19
5	8	6,79	1,21
6,2	6	7,31	-1,31
6,5	8	7,44	0,56

$$y = 4.64 + 0.43x$$



# Regresión y correlación

- Sobre los residuos: Lo que queda “sin explicar”

$$e_i = y_i - \hat{y}_i$$

- El error residual medio es nulo:

$$\bar{e} = \frac{1}{n} \sum_{i=1}^n e_i = \frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_i) = \dots = 0$$

# Regresión y correlación

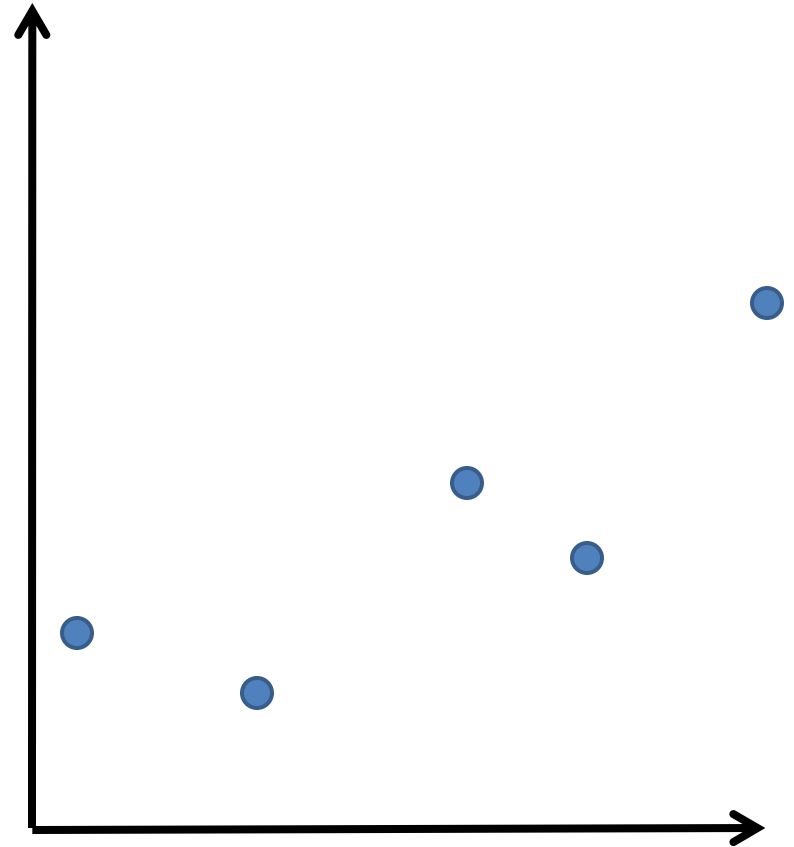
- La varianza de los residuos (varianza residual) es mínima:

$$S_e^2 = \frac{1}{n} \sum_{i=1}^n e_i^2 = \frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_i)^2 = \frac{1}{n} \sum_{i=1}^n (y_i - \hat{a} - \hat{b}x_i)^2$$

- Por definición del ajuste por mínimos cuadrados...

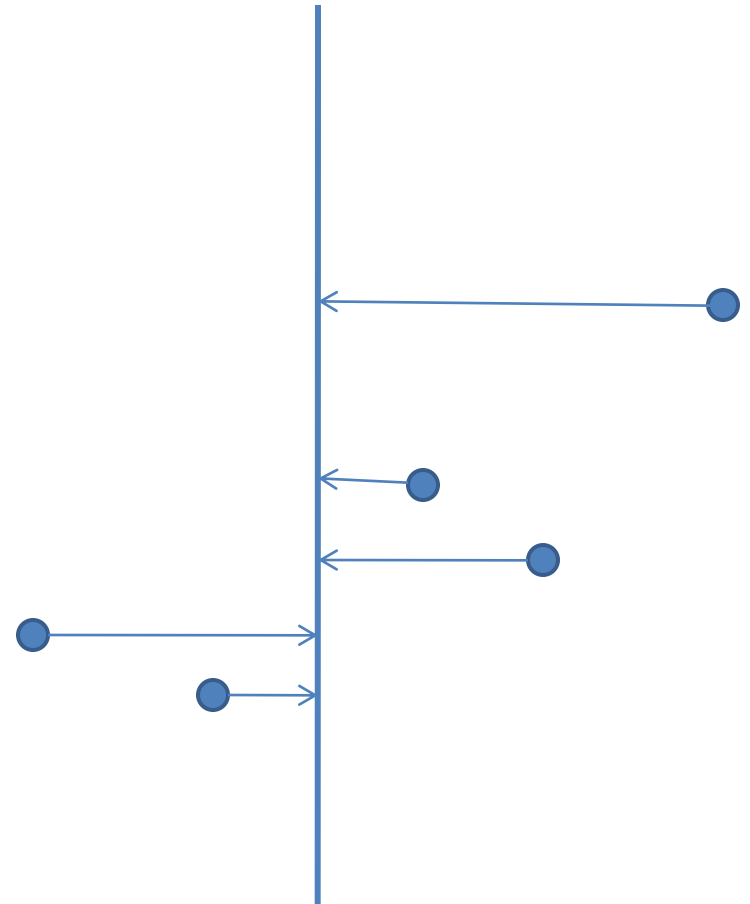
# Regresión y correlación

- Como medimos la “bondad del ajuste”?
  - ▣ Vamos tener una relación entre la variabilidad de  $Y$  e la variabilidad del residuo  $e$ .



# Regresión y correlación

- Variabilidad de Y:  
proyectamos los valores  
observados sobre el eje OY.

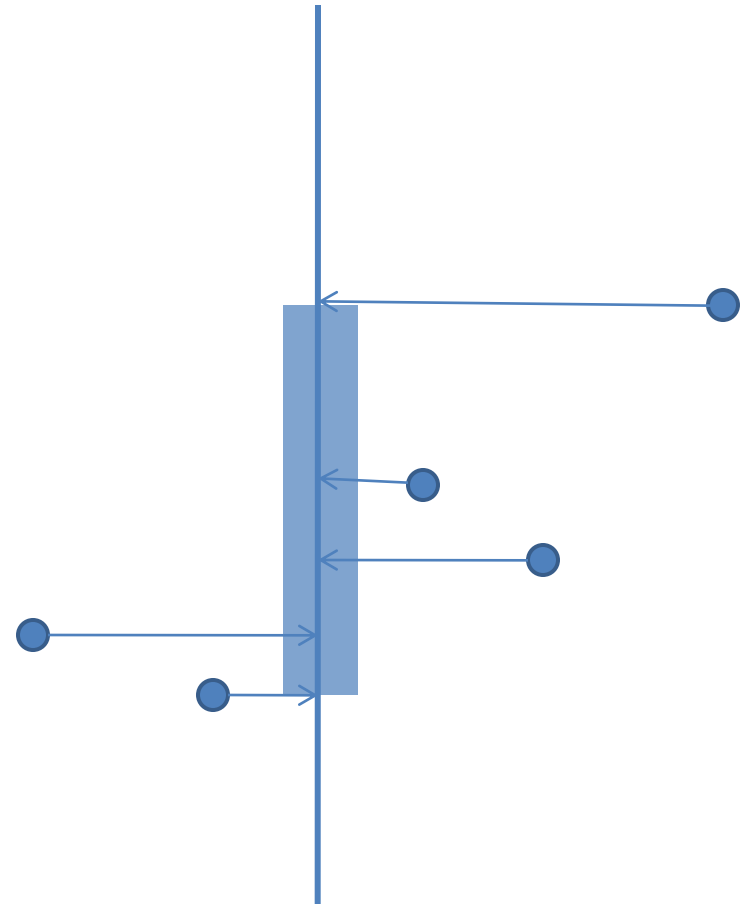


# Regresión y correlación

- Variabilidad de Y:  
proyectamos los valores  
observados sobre el eje OY.

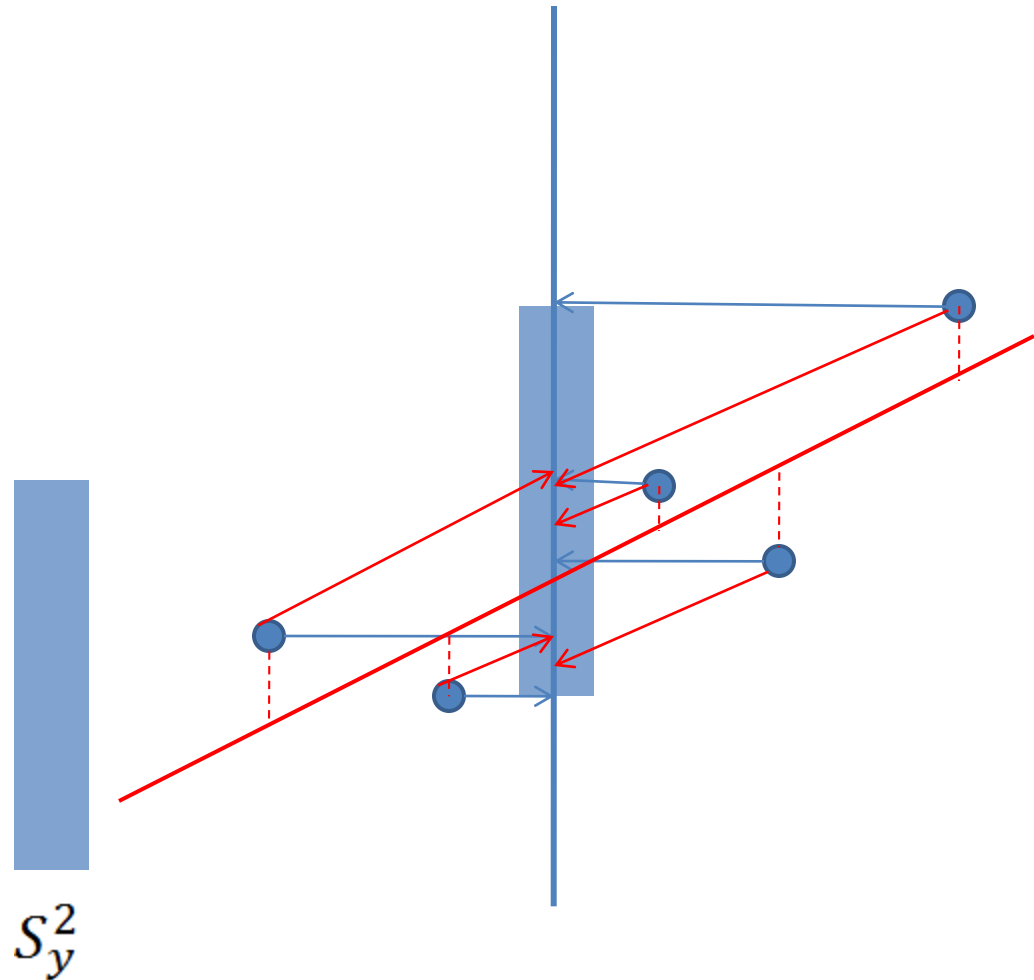


$S_y^2$



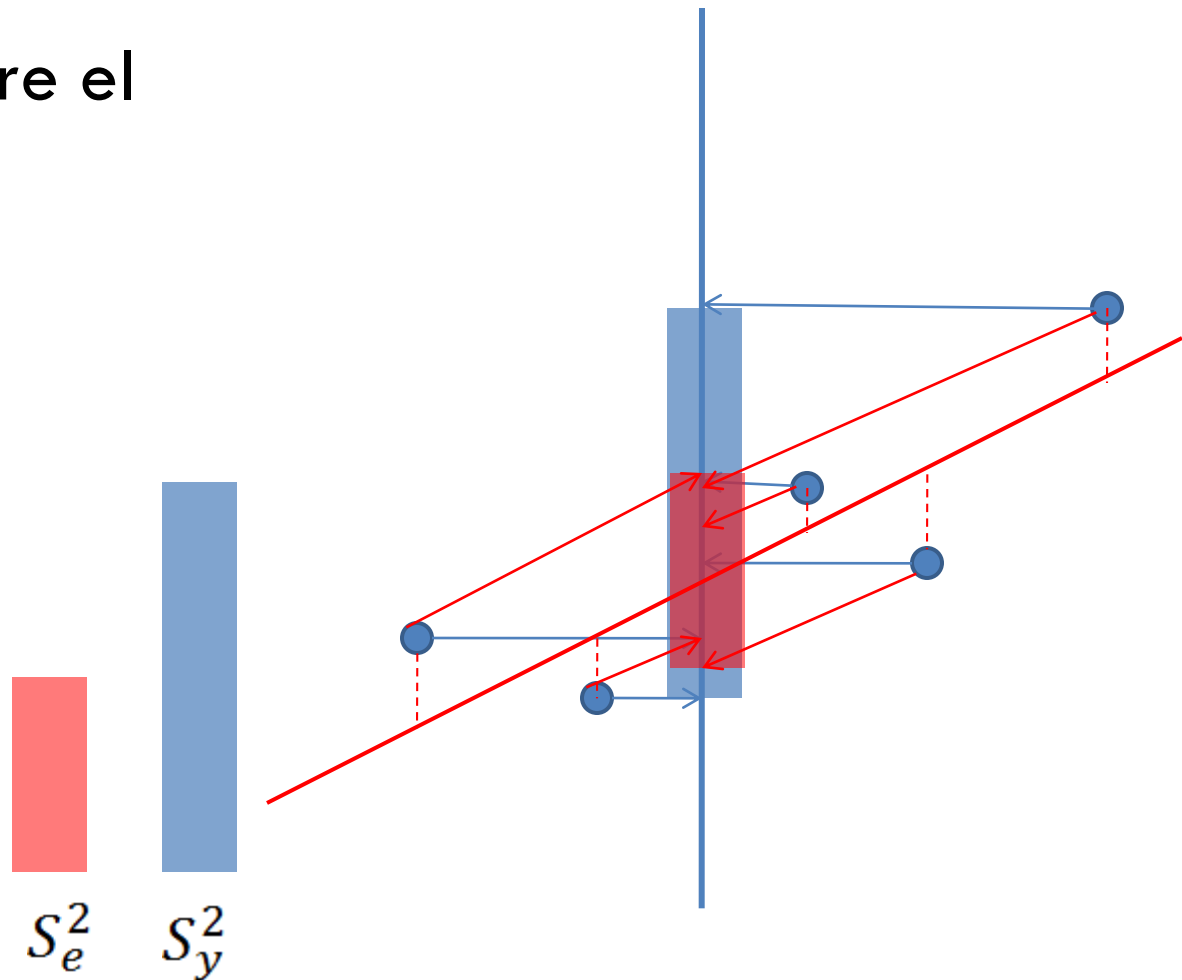
# Regresión y correlación

- Residuos:  
proyectamos sobre el  
eje OY.



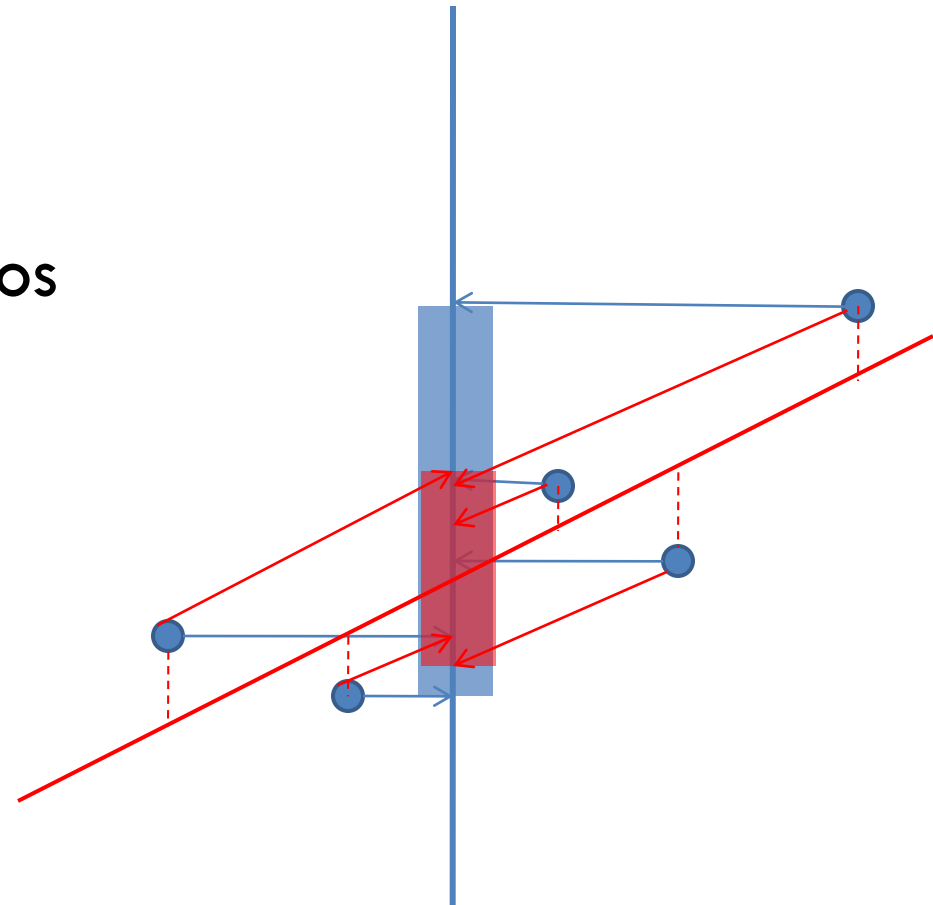
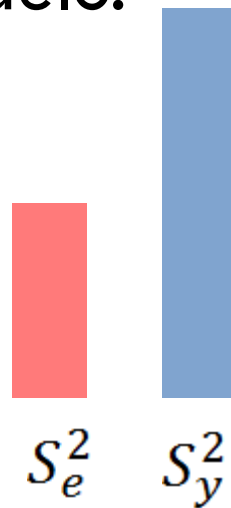
# Regresión y correlación

- Residuos:  
proyectamos sobre el  
eje OY.



# Regresión y correlación

- Residuos:  
proyectamos sobre el  
eje OY.
- La varianza de los residuos nos  
da la variabilidad (de Y) no  
explicada por el modelo.



# Regresión y correlación

## □ Bondad do ajuste:

- La dispersión del error residual será una fracción de la dispersión de Y
- Cuanto menor sea esta dispersión mejor será o ajuste

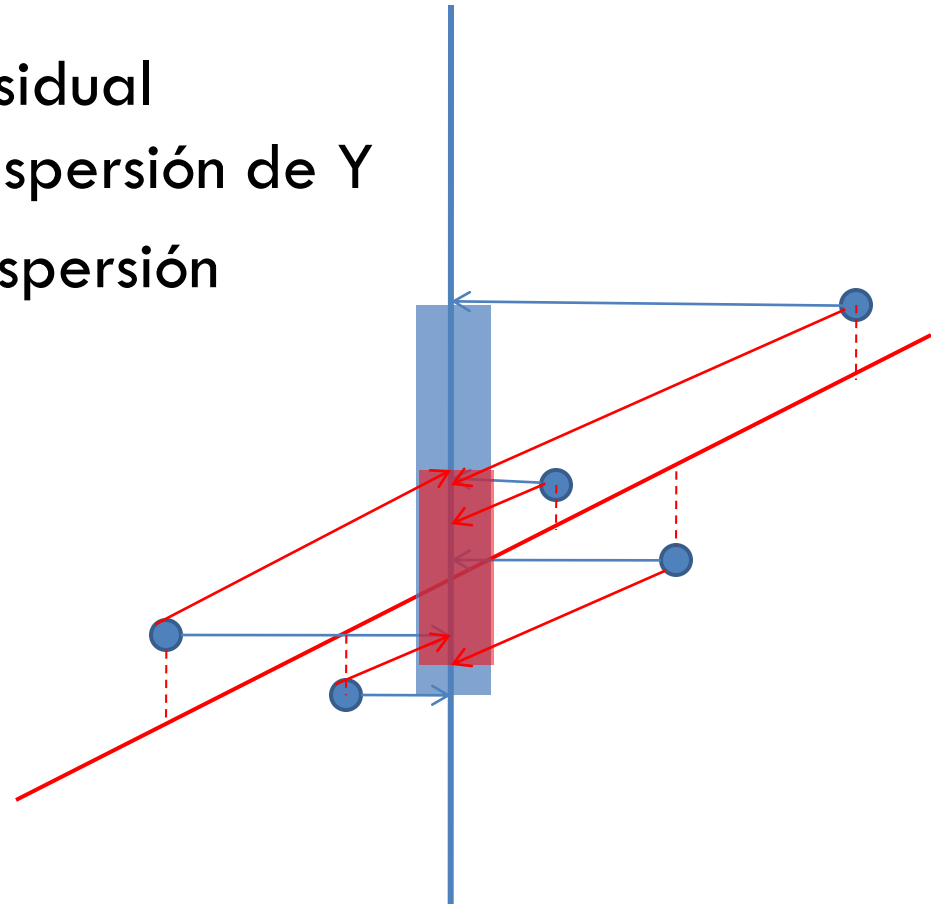
$$R^2 = 1 - \frac{S_e^2}{S_y^2}$$



$S_e^2$



$S_y^2$



# Regresión y correlación

## □ El coeficiente de determinación:

$$R^2 = 1 - \frac{S_e^2}{S_y^2} \equiv r^2 = \left( \frac{S_{xy}}{S_x S_y} \right)^2$$

proporciona una medida de la bondad del ajuste de una recta de regresión (o de otro ajuste paramétrico).

# Regresión y correlación

- una vez que tenemos la relación:

$$y = a + bx$$

$$b = r \frac{S_y}{S_x}$$

$$a = \bar{y} - b\bar{x}$$

- podemos hacer predicciones (!): dado un valor  $x_0$  ,  
calcular:

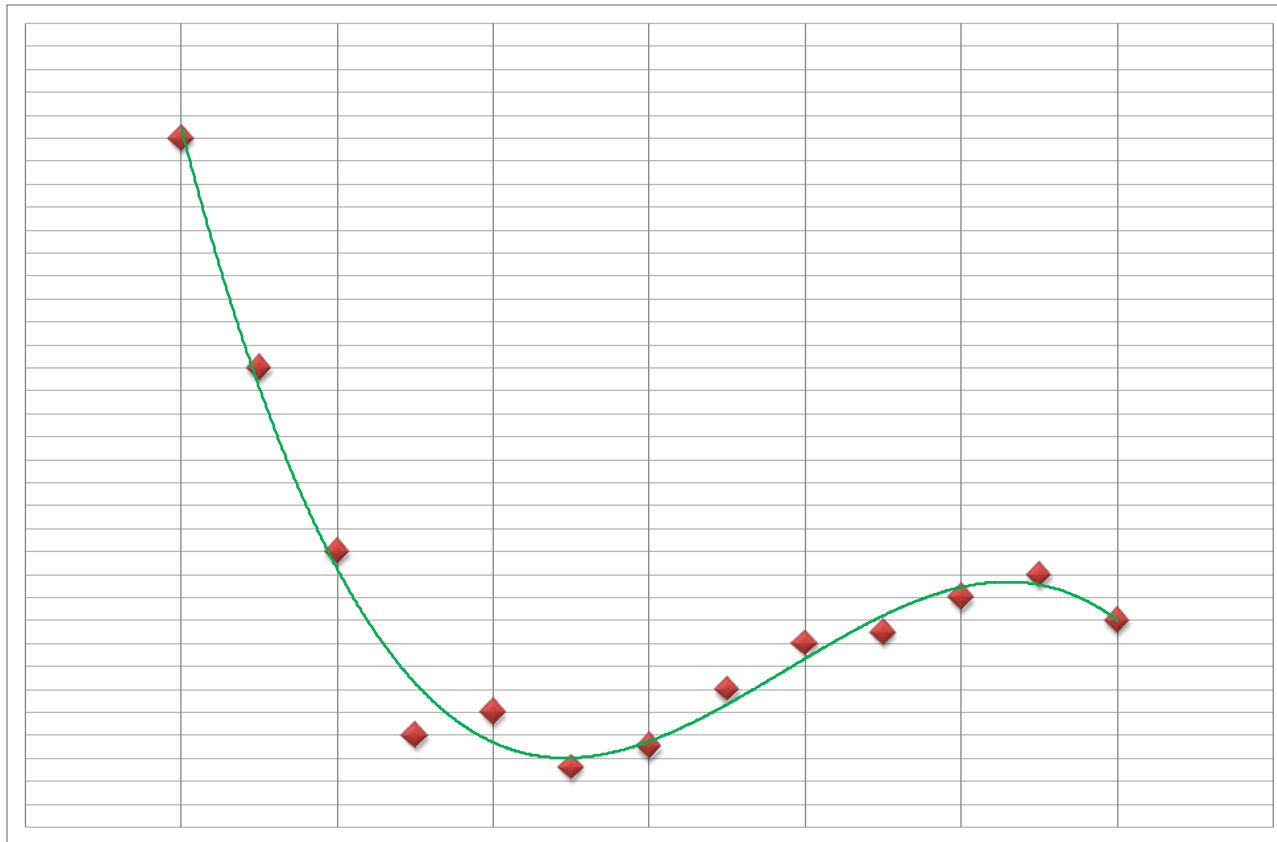
$$y_0 = a + bx_0$$

# Regresión y correlación

- Algunas aclaraciones:
  - ▣ La recta de  $Y$  sobre  $X$  no es la misma que la de  $X$  sobre  $Y$ ...
  - ▣ El método es sensible a datos atípicos
  - ▣ No todo es lineal...
    - Pero también hay mínimos cuadrados!
  - ▣ Non todo es paramétrico...
    - Pero precisaríamos de otro curso...

# Regresión y correlación

- ejemplos de relaciones NO LINEALES
  - ▣ Polinomio de grado 3



# Regresión y correlación

- ejemplos de relaciones NO LINEALES
  - Curvas de ingresos-gastos
  - Curva de consumo de vino, de riesgo de infarto, de desgaste, de supervivencia,...

# Regresión y correlación

- Cuatro conjuntos de datos (Anscombe):

<b>x1</b>	<b>y1</b>	<b>x2</b>	<b>y2</b>	<b>x3</b>	<b>y3</b>	<b>x4</b>	<b>y4</b>
10	8,04	10	9,14	10	7,46	8	6,58
8	6,95	8	8,14	8	6,77	8	5,76
13	7,58	13	8,74	13	12,74	8	7,71
9	8,81	9	8,77	9	7,11	8	8,84
11	8,33	11	9,26	11	7,81	8	8,47
14	9,96	14	8,1	14	8,84	8	7,04
6	7,24	6	6,13	6	6,08	8	5,25
4	4,26	4	3,1	4	5,39	19	12,5
12	10,84	12	9,13	12	8,15	8	5,56
7	4,82	7	7,26	7	6,42	8	7,91
5	5,68	5	4,74	5	5,73	8	6,89

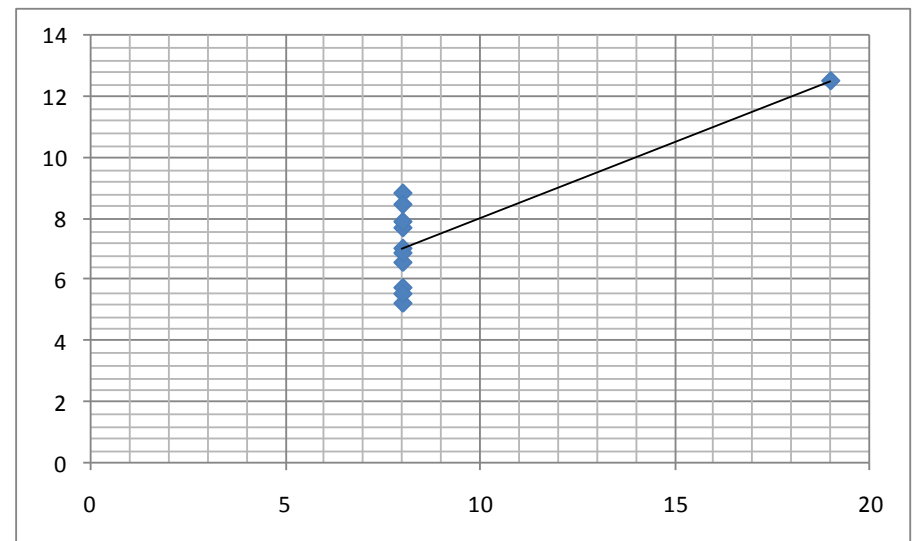
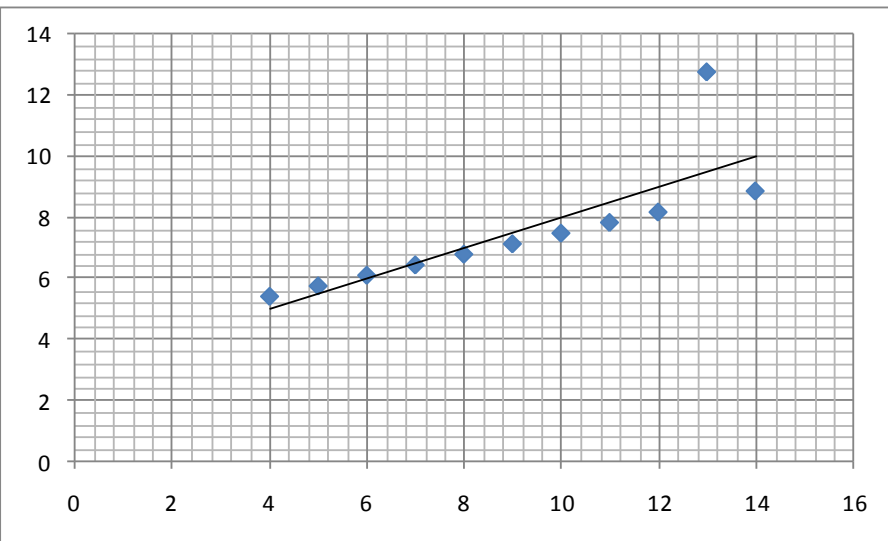
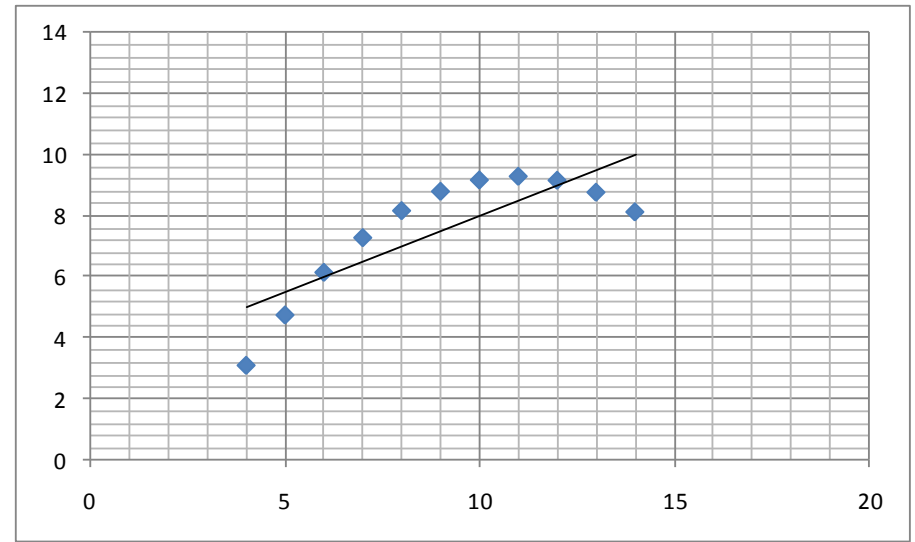
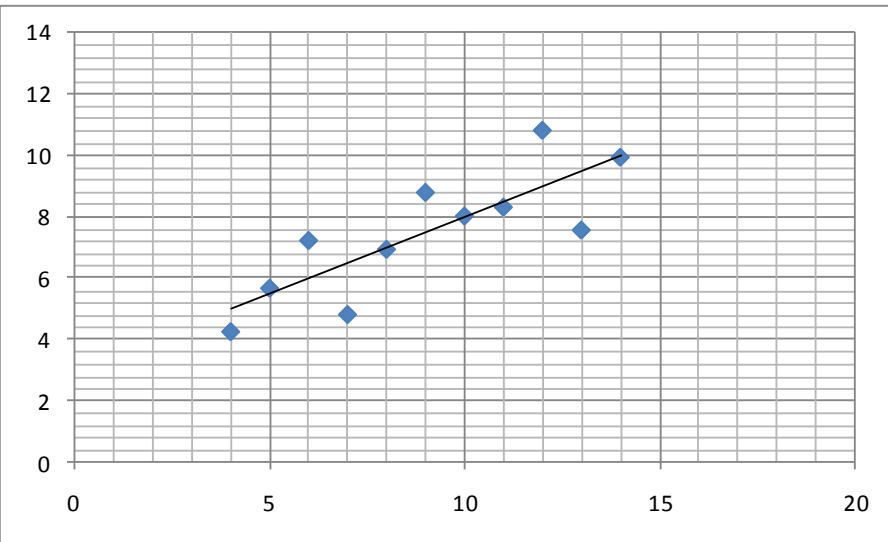
# Regresión y correlación

- Cuatro conjuntos de datos: CUARTETO DE ANSCOMBE

x1	y1	x2	y2	x3	y3	x4	y4
10	8,04	10	9,14	10	7,46	8	6,58
8	6,95	8	8,14	8	6,77	8	5,76
13	7,58	13	8,74	13	12,74	8	7,71
9	8,81	9	8,77	9	7,11	8	8,84
11	8,33	11	9,26	11	7,81	8	8,47
14	9,96	14	8,1	14	8,84	8	7,04
6	7,24	6	6,13	6	6,08	8	5,25
4	4,26	4	3,1	4	5,39	19	12,5
12	10,84	12	9,13	12	8,15	8	5,56
7	4,82	7	7,26	7	6,42	8	7,91
5	5,68	5	4,74	5	5,73	8	6,89

Media X	9
Media Y	7,5
Var X	11
Var Y	4,13
Recta	$y=3+0.5x$

# Regresión y correlación



# Portal educativo del IGE

- <http://www.ige.eu/estatico/educacion/index.htm>
- Varios módulos de contenidos:
  - Estadística descriptiva
  - Probabilidad
  - Regresión
  - Series de tiempo

# Applet interactivo

- APPLETS:

- <http://www.shodor.org/interactivate/>

- Para:

- Estudiantes

- Profesores

# Módulo 2. Estadística descriptiva de dos variables



Salvador Naya Fernández  
salvador.naya@udc.es

# Modulo 3: Azar y Probabilidad

Salvador Naya

Introducción

Experimentos  
y sucesos

Definición de  
probabilidad

Probabilidad  
condicionada  
e  
independencia  
de sucesos

Teorema de  
Bayes

Variables  
aleatorias

Noción de  
variable  
aleatoria

Variable  
aleatoria  
discreta

Variable

Cursillo de Estadística. Medellín 2010

Azar y probabilidad

Profesor: Salvador Naya

Departamento de Matemáticas. Universidade da Coruña

# Introducción

Salvador Naya

## Introducción

Experimentos  
y sucesos

Definición de  
probabilidad

Probabilidad  
condicionada  
e  
independencia  
de sucesos

Teorema de  
Bayes

Variables  
aleatorias

Noción de  
variable  
aleatoria

Variable  
aleatoria  
discreta

Variable

- El concepto de probabilidad está asociado a experimentos (procesos de observación) donde existe incertidumbre sobre el resultado final, que desde un punto de vista práctico son la mayoría de los experimentos reales.
- La Teoría de la Probabilidad es importante como soporte teórico de la Estadística (Inferencia Estadística) y como herramienta en el estudio de la mayoría de las áreas de conocimiento: Ingeniería, Economía, Sociología, Medicina, Biología, etc.
- El origen de la Teoría de la Probabilidad está ligado al estudio de los juegos de azar, siendo pioneros los trabajos realizados por G. Cardano y G. Galilei en el siglo XVI. Actualmente constituye un área científica de intensa investigación.

# Experimentos y sucesos

Salvador Naya

Introducción

Experimentos  
y sucesos

Definición de  
probabilidad

Probabilidad  
condicionada  
e  
independencia  
de sucesos

Teorema de  
Bayes

Variables  
aleatorias

Noción de  
variable  
aleatoria

Variable  
aleatoria  
discreta

Variable

## Experimentos

- Un **experimento** es “un proceso por medio del cual se obtiene una observación”.
- Un **experimento determinista** es el que al realizarse repetidas veces, en idénticas condiciones, proporciona siempre el mismo resultado y, por tanto, puede predecirse de antemano.
- Un **experimento aleatorio** es el que puede dar lugar a diferentes resultados, conocidos previamente, sin que sea posible predecir cuál va a ser el resultado que va a ocurrir en una determinada realización del experimento.
- La Teoría de la Probabilidad y la Estadística estudian los experimentos aleatorios que, en mayor o menor medida, son todos los experimentos reales.

# Experimentos y sucesos

Salvador Naya

Introducción

Experimentos  
y sucesos

Definición de  
probabilidad

Probabilidad  
condicionada  
e  
independencia  
de sucesos

Teorema de  
Bayes

Variables  
aleatorias

Noción de  
variable  
aleatoria

Variable  
aleatoria  
discreta

Variable

## Álgebra de sucesos

- **Suceso elemental o simple:** es cada uno de los posibles resultados del experimento aleatorio.
- **Espacio muestral:** es el conjunto formado por todos los sucesos elementales. Lo denotaremos por  $\Omega = \{\omega/\omega \text{ es un suceso elemental}\}$ . Se clasifica en: discreto (si es finito o infinito numerable) y continuo.
- **Suceso:** es un subconjunto del espacio muestral. Son sucesos de interés:  $\Omega$ , el **suceso seguro**, formado por todos los sucesos elementales y  $\emptyset$ , el **suceso imposible**, que no contiene elementos.
- **Álgebra de sucesos:** es el conjunto formado por todos los sucesos asociados a un experimento aleatorio. Lo denotaremos por  $\mathcal{A} = \{A/A \text{ es un suceso}\}$ .

# Experimentos y sucesos

Salvador Naya

Introducción

**Experimentos y sucesos**

Definición de probabilidad

Probabilidad condicionada e independencia de sucesos

Teorema de Bayes

Variables aleatorias

Noción de variable aleatoria

Variable aleatoria discreta

Variable

## Ejemplo 1

Considérese el experimento aleatorio “lanzar un dado y observar el número de puntos obtenido”. Los sucesos elementales son  $\omega_i =$  “se obtienen  $i$  puntos”, donde  $i = 1, 2, \dots, 6$ . Son sucesos  $A =$  “se obtiene un número par” = “el resultado es 2, 4 o 6” y  $B =$  “se obtiene un número mayor que 2” = “el resultado es 3, 4, 5 o 6”.

## Ejemplo 2

Considérese el experimento aleatorio “tiempo de ejecución de un programa”. Los sucesos elementales son  $\omega_t =$  “la ejecución ha durado  $t$  segundos”, con  $t \in \mathbb{R}$ ,  $t \geq 0$ . Son sucesos  $C =$  “el tiempo de ejecución es superior a 10 segundos” y  $D =$  “el tiempo de ejecución está entre 5 y 15 segundos”.

# Experimentos y sucesos

Salvador Naya

Introducción

Experimentos  
y sucesos

Definición de  
probabilidad

Probabilidad  
condicionada  
e  
independencia  
de sucesos

Teorema de  
Bayes

Variables  
aleatorias

Noción de  
variable  
aleatoria

Variable  
aleatoria  
discreta

Variable

## Álgebra de sucesos. Operaciones

- **Unión de sucesos:** si  $A, B \in \mathcal{A}$ , se define el suceso unión,  $A \cup B$ , como el que ocurre si sucede  $A$  o sucede  $B$ .
- **Intersección de sucesos:** si  $A, B \in \mathcal{A}$ , se define el suceso intersección,  $A \cap B$ , como el que ocurre si sucede  $A$  y sucede  $B$ . Por sencillez,  $A \cap B$  también se escribe  $AB$ .
- **Suceso complementario o contrario:** si  $A \in \mathcal{A}$ , se define el suceso contrario,  $\bar{A}$ , como el que ocurre si no sucede  $A$ .
- **Inclusión de sucesos:** si  $A, B \in \mathcal{A}$ , se dice que  $A$  está contenido en  $B$  o que  $A$  implica  $B$ ,  $A \subset B$ , si siempre que sucede  $A$  ocurre  $B$ .

## Álgebra de sucesos. Operaciones

- **Diferencia de sucesos:** si  $A, B \in \mathcal{A}$ , se define el suceso diferencia,  $A \setminus B$ , como el que ocurre si sucede  $A$  y no sucede  $B$ , esto es,  
$$A \setminus B = A \cap \bar{B}$$
- **Diferencia simétrica de sucesos:** si  $A, B \in \mathcal{A}$ , se define el suceso diferencia simétrica,  $A \nabla B$ , como el que ocurre si sucede sólo  $A$  o sólo  $B$ , esto es,  
$$A \nabla B = (A \cup B) \setminus (A \cap B) = (A \cap \bar{B}) \cup (B \cap \bar{A})$$
- **Sucesos incompatibles:** dos sucesos  $A, B \in \mathcal{A}$  son incompatibles si  $A \cap B = \emptyset$ .

## Álgebra de sucesos. Operaciones

- **Conjunto exhaustivo de sucesos:**  $\{A_1, A_2, \dots, A_n / A_i \in \mathcal{A}\}$  es un conjunto exhaustivo de sucesos si  $\bigcup_{i=1}^n A_i = \Omega$
- **Conjunto completo de sucesos:**  $\{A_1, A_2, \dots, A_n / A_i \in \mathcal{A}\}$  es un conjunto completo de sucesos si es exhaustivo y los sucesos son incompatibles dos a dos:  
 $\bigcup_{i=1}^n A_i = \Omega$  y  $A_i \cap A_j = \emptyset$  si  $i \neq j$   
A un “conjunto completo de sucesos” también se le denomina **partición** del espacio muestral. El conjunto de los sucesos elementales es una clase completa de sucesos y la partición más fina del espacio muestral.
- El álgebra de sucesos,  $\mathcal{A}$ , asociada a un experimento aleatorio tiene estructura de álgebra de Boole respecto a las operaciones unión e intersección

# Experimentos y sucesos

Salvador Naya

Introducción

Experimentos  
y sucesos

Definición de  
probabilidad

Probabilidad  
condicionada  
e  
independencia  
de sucesos

Teorema de  
Bayes

Variables  
aleatorias

Noción de  
variable  
aleatoria

Variable  
aleatoria  
discreta

Variable

## Álgebra de sucesos. Propiedades

- **Conmutativa.**  $A \cup B = B \cup A$ ,  $A \cap B = B \cap A$
- **Asociativa.]**  $A \cup (B \cup C) = (A \cup B) \cup C$ ,  
 $A \cap (B \cap C) = (A \cap B) \cap C$
- **Elemento neutro.** El suceso imposible ( $\emptyset$ ) para la unión  $A \cup \emptyset = A$  y el suceso seguro ( $\Omega$ ) para la intersección ( $A \cap \Omega = A$ ).
- **Complementario.** Dado  $A \in \mathcal{A}$  existe  $\bar{A}$ , que llamaremos suceso complementario o contrario de  $A$ , tal que  $A \cup \bar{A} = \Omega$  y  $A \cap \bar{A} = \emptyset$ .

# Experimentos y sucesos

Salvador Naya

Introducción

Experimentos  
y sucesos

Definición de  
probabilidad

Probabilidad  
condicionada  
e  
independencia  
de sucesos

Teorema de  
Bayes

Variables  
aleatorias

Noción de  
variable  
aleatoria

Variable  
aleatoria  
discreta

Variable

## Álgebra de sucesos. Propiedades

- **Idempotente.**  $A \cup A = A$ ,  $A \cap A = A$
- **Simplificativa.**  $A \cup (A \cap B) = A$ ,  $A \cap (A \cup B) = A$
- **Relativas al elemento neutro.**  $A \cup \Omega = \Omega$ ,  $A \cap \emptyset = \emptyset$
- **Leyes de De Morgan.**  $\overline{A \cup B} = \bar{A} \cap \bar{B}$ ,  $\overline{A \cap B} = \bar{A} \cup \bar{B}$

# Experimentos y sucesos

Salvador Naya

Introducción

**Experimentos  
y sucesos**

Definición de  
probabilidad

Probabilidad  
condicionada  
e  
independencia  
de sucesos

Teorema de  
Bayes

Variables  
aleatorias

Noción de  
variable  
aleatoria

Variable  
aleatoria  
discreta

## Ejemplo 3

Respecto al experimento del ejemplo 1 se obtienen los siguientes sucesos:

$A \cup B =$  “obtener 2, 3, 4, 5 o 6”.

$A \cap B =$  “obtener 4 o 6”.

$\bar{A} =$  “obtener un número impar”.

$\bar{B} =$  “obtener 1 o 2”.

$A \setminus B =$  “obtener el 2”.

$B \setminus A =$  “obtener 3 o 5”.

$A \nabla B =$  “obtener 2, 3 o 5”.

# Experimentos y sucesos

Salvador Naya

Introducción

**Experimentos  
y sucesos**

Definición de  
probabilidad

Probabilidad  
condicionada  
e  
independencia  
de sucesos

Teorema de  
Bayes

Variables  
aleatorias

Noción de  
variable  
aleatoria

Variable  
aleatoria  
discreta

Variable

## Ejemplo 4

Respecto al experimento del ejemplo 2 se obtienen los siguientes sucesos:

$C \cup D =$  “el tiempo de ejecución es superior a 5 segundos”.

$C \cap D =$  “el tiempo de ejecución está entre 10 y 15 segundos”.

$\bar{C} =$  “el tiempo de ejecución es inferior o igual a 10 segundos”.

$\bar{D} =$  “el tiempo de ejecución es menor o igual que 5 segundos o mayor o igual que 15 segundos”.

$C \setminus D =$  “el tiempo de ejecución es mayor o igual que 15 segundos”.

$D \setminus C =$  “el tiempo de ejecución es superior a 5 segundos y menor o igual que 10 segundos”.

$C \nabla D =$  “el tiempo de ejecución está en  $(5, 10] \cup [15, \infty)$ ”.

# Definición de probabilidad

## Definición axiomática de Kolmogorov

La **probabilidad** ( $P$ ) asociada a un experimento aleatorio es una aplicación del álgebra de sucesos ( $\mathcal{A}$ ) en  $\mathbb{R}$

$$P : \mathcal{A} \longrightarrow \mathbb{R}$$

verificando los siguientes axiomas:

- 1 Para todo suceso  $A$ ,  $P(A) \geq 0$
- 2  $P(\Omega) = 1$
- 3 ( $\sigma$ -aditividad) Si  $\{A_n\}_{n=1}^{\infty}$  es una sucesión de sucesos incompatibles dos a dos, entonces:

$$P\left(\bigcup_{n=1}^{\infty} A_n\right) = \sum_{n=1}^{\infty} P(A_n)$$

Salvador Naya

Introducción

Experimentos  
y sucesos

**Definición de  
probabilidad**

Probabilidad  
condicionada  
e  
independencia  
de sucesos

Teorema de  
Bayes

Variables  
aleatorias

Noción de  
variable  
aleatoria

Variable  
aleatoria  
discreta

Variable

# Definición de probabilidad

Salvador Naya

Introducción

Experimentos  
y sucesos

**Definición de  
probabilidad**

Probabilidad  
condicionada  
e  
independencia  
de sucesos

Teorema de  
Bayes

Variables  
aleatorias

Noción de  
variable  
aleatoria

Variable  
aleatoria  
discreta

Variable

## Definición

Llamaremos **espacio de probabilidad** a la terna,  $(\Omega, \mathcal{A}, \mathcal{P})$  formada por el espacio muestral  $(\Omega)$ , el álgebra de sucesos  $(\mathcal{A})$  y la aplicación  $(\mathcal{P})$  verificando los anteriores axiomas.

# Definición de probabilidad

Salvador Naya

Introducción

Experimentos  
y sucesos

**Definición de  
probabilidad**

Probabilidad  
condicionada  
e  
independencia  
de sucesos

Teorema de  
Bayes

Variables  
aleatorias

Noción de  
variable  
aleatoria

Variable  
aleatoria  
discreta

Variable

## Ejemplo 5

En relación con el experimento del ejemplo 1, puede definirse la función de probabilidad a partir de la probabilidad de los sucesos elementales,  $A_i =$  “obtener el número  $i$ ”, de la siguiente forma:  $P(A_i) = \frac{1}{6}, i = 1, 2, \dots, 6$

## Ejemplo 6

En relación con el experimento del ejemplo 2, puede definirse la función de probabilidad a partir de la probabilidad de sucesos de la forma  $A_t =$  “la duración de la ejecución del programa es inferior a  $t$  segundos”, como  $P(A_t) = 1 - e^{-t} (t > 0)$ .

# Definición de probabilidad

Salvador Naya

Introducción

Experimentos  
y sucesos

Definición de  
probabilidad

Probabilidad  
condicionada  
e  
independencia  
de sucesos

Teorema de  
Bayes

Variables  
aleatorias

Noción de  
variable  
aleatoria

Variable  
aleatoria  
discreta

Variable

## Propiedades

- 1  $P(\emptyset) = 0$ .
- 2 Si  $\{A_i\}_{i=1}^n$  es un conjunto de sucesos incompatibles dos a dos entonces,  $P\left(\bigcup_{j=1}^n A_j\right) = \sum_{j=1}^n P(A_j)$
- 3  $P(\bar{A}) = 1 - P(A)$ .
- 4 Para cualquier suceso  $A$ ,  $0 \leq P(A) \leq 1$ .
- 5 Si  $A \subset B$  entonces  $P(A) \leq P(B)$  y  $P(B \setminus A) = P(B) - P(A)$ .
- 6 Para dos sucesos cualesquiera  $A$  y  $B$  se verifica que,  $P(A \cup B) = P(A) + P(B) - P(A \cap B)$ .

# Definición de probabilidad

Salvador Naya

Introducción

Experimentos  
y sucesos

Definición de  
probabilidad

Probabilidad  
condicionada  
e  
independencia  
de sucesos

Teorema de  
Bayes

Variables  
aleatorias

Noción de  
variable  
aleatoria

Variable  
aleatoria  
discreta

## Ejemplo 7

La probabilidad de que el estudiante A apruebe un examen es 0'5, la probabilidad de que apruebe B es 0'3 y la probabilidad de que aprueben los dos es 0'2.

- La probabilidad de que al menos uno de los dos apruebe es  $P(A \cup B) = 0'5 + 0'3 - 0'2 = 0'6$ .
- La probabilidad de que exactamente uno de los dos apruebe es  $P(A \nabla B) = P(A) + P(B) - 2P(A \cap B) = 0'4$ .
- La probabilidad de que no apruebe ni A ni B es  $P(\bar{A} \cap \bar{B}) = 1 - P(A \cup B) = 1 - 0'6 = 0'4$ .
- La probabilidad de que apruebe A pero no B es  $P(A \cap \bar{B}) = P(A \setminus B) = P(A) - P(A \cap B) = 0'5 - 0'2 = 0'3$ .

# Definición de probabilidad

Salvador Naya

Introducción

Experimentos  
y sucesos

Definición de  
probabilidad

Probabilidad  
condicionada  
e  
independencia  
de sucesos

Teorema de  
Bayes

Variables  
aleatorias

Noción de  
variable  
aleatoria

Variable  
aleatoria  
discreta

Variable

## Ejemplo 8

Supóngase que la probabilidad de obtener el número  $i$  al lanzar un dado es inversamente proporcional a dicho número. Calcular la probabilidad de obtener un número par en una tirada.

Llamamos  $p_i = P(\text{"obtener el número } i\text{"}) = k/i$ ,  $i = 1, 2, \dots, 6$ , con  $k$  una constante por determinar, que obtenemos de la siguiente igualdad

$$\sum_{i=1}^6 p_i = k \sum_{i=1}^6 \frac{1}{i} = P(\Omega) = 1 \implies k = \frac{60}{147}$$

Por tanto,  $P(\text{"obtener un número par"}) = \frac{55}{147}$ .

# Definición de probabilidad

Salvador Naya

Introducción

Experimentos  
y sucesos

Definición de  
probabilidad

Probabilidad  
condicionada  
e  
independencia  
de sucesos

Teorema de  
Bayes

Variables  
aleatorias

Noción de  
variable  
aleatoria

Variable  
aleatoria  
discreta

Variable

## Asignación de probabilidades

- **Método de las frecuencias:** Definir la probabilidad del suceso como el límite de las frecuencias relativas.
- **Método clásico:** En los espacios muestrales finitos **equiprobables**, podemos calcular la probabilidad del suceso  **$A$**  como el cociente entre el número de “*casos favorables*” en que sucede  **$A$**  y el número de “*casos posibles*” que se pueden dar. Esta regla se conoce como definición clásica o **Ley de Laplace**.
- **Método subjetivo:** en el que una determinada persona asigna de forma subjetiva probabilidades a cada uno de los posibles resultados de un proceso según su propio juicio sobre la verosimilitud de cada resultado.

# Probabilidad condicionada e independencia de sucesos

Salvador Naya

Introducción

Experimentos y sucesos

Definición de probabilidad

Probabilidad condicionada e independencia de sucesos

Teorema de Bayes

Variables aleatorias

Noción de variable aleatoria

Variable aleatoria discreta

Variable

## Ejemplo 9

En un curso de Estadística de 80 estudiantes aprobaron 50, de los que 35 eran chicas. La probabilidad de que haya aprobado un alumno elegido al azar es:  $P(\text{aprobar}) = \frac{50}{80} = 0'625$   
Pero si el número de chicas que participaron en el curso fue de 45, entonces la probabilidad de que haya aprobado un alumno elegido al azar *sabiendo que es una chica*, es:

$$\begin{aligned} P(\text{aprobar/ser chica}) &= \frac{P(\text{aprobar y ser chica})}{P(\text{ser chica})} \\ &= \frac{35/80}{45/80} = 0'777 \end{aligned}$$

# Probabilidad condicionada e independencia de sucesos

Salvador Naya

Introducción

Experimentos y sucesos

Definición de probabilidad

Probabilidad condicionada e independencia de sucesos

Teorema de Bayes

Variables aleatorias

Noción de variable aleatoria

Variable aleatoria discreta

Variable

## Definición

Sean  $A$  y  $B$  dos sucesos cualesquiera con  $P(B) > 0$ . Se define la **probabilidad del suceso  $A$  condicionada al suceso  $B$**  y se representa por  $P(A/B)$  como:

$$P(A/B) = \frac{P(A \cap B)}{P(B)}$$

# Probabilidad condicionada e independencia de sucesos

Salvador Naya

Introducción

Experimentos  
y sucesos

Definición de  
probabilidad

Probabilidad  
condicionada  
e  
independencia  
de sucesos

Teorema de  
Bayes

Variables  
aleatorias

Noción de  
variable  
aleatoria

Variable  
aleatoria  
discreta

Variable

## Probabilidad condicionada. Comentarios

- 1 La probabilidad condicionada es muy importante en la práctica, ya que, en muchas situaciones, pequeñas modificaciones en la información básica producen cambios sustanciales en las probabilidades condicionadas.
- 2 Con la definición anterior, es fácil probar que la probabilidad condicionada a un suceso  $B$  verifica la axiomática de la probabilidad dada en la definición.
- 3 Es importante diferenciar entre  $P(AB)$  y  $P(A/B)$ : la primera indica la probabilidad de ocurrencia de  $A$  y  $B$  conjuntamente, por tanto siempre es menor o igual que  $P(A)$ ; y la segunda indica la probabilidad de ocurrencia de  $A$  cuando es conocido que ha ocurrido el suceso  $B$  y puede ser menor, igual o mayor que  $P(A)$ .

# Probabilidad condicionada e independencia de sucesos

Salvador Naya

Introducción

Experimentos  
y sucesos

Definición de  
probabilidad

Probabilidad  
condicionada  
e  
independencia  
de sucesos

Teorema de  
Bayes

Variables  
aleatorias

Noción de  
variable  
aleatoria

Variable  
aleatoria  
discreta

## Ejemplo 10

En un almacén se dispone de diez motores de los cuales tres son defectuosos. Si se eligen dos motores al azar y Denominando por  $D_i$  al suceso “el motor elegido en lugar  $i$ -ésimo es defectuoso” y  $N_i$  al suceso “el motor elegido en lugar  $i$ -ésimo es no defectuoso”, se pueden calcular las siguientes probabilidades condicionadas

$$\textcircled{1} \quad P(D_2/N_1) = \frac{P(N_1 \cap D_2)}{P(N_1)} = \frac{7/10 \cdot 3/9}{7/10} = \frac{3}{9}$$

$$\textcircled{2} \quad P(D_2/D_1) = \frac{P(D_1 \cap D_2)}{P(D_1)} = \frac{3/10 \cdot 2/9}{3/10} = \frac{2}{9}$$

$$\textcircled{3} \quad P(D_2) = \frac{3}{10}$$

# Probabilidad condicionada e independencia de sucesos

Salvador Naya

Introducción

Experimentos  
y sucesos

Definición de  
probabilidad

Probabilidad  
condicionada  
e  
independencia  
de sucesos

Teorema de  
Bayes

Variables  
aleatorias

Noción de  
variable  
aleatoria

Variable  
aleatoria  
discreta

Variable

## Ejemplo 11

En una encuesta realizada en Medellín se ha determinado que el 40 % de los encuestados lee el periódico *El Colombiano*, el 15 % lee *La República* y el 3 % lee ambos periódicos.

- 1 Seleccionado al azar un lector de *La República*, calcular la probabilidad de que lea *El Colombiano*.

Sea  $C$  el suceso "lee *El Colombiano*.", y  $R$  el suceso "lee *La República*", entonces  $P(C/R) = \frac{P(C \cap R)}{P(R)} = \frac{3}{15} = 0'2$

- 2 Si se ha elegido un lector de *El Colombiano*, calcular la probabilidad de que no lea *La República*.

$$P(\bar{R}/C) = 1 - \frac{P(R \cap C)}{P(C)} = 1 - \frac{3}{40} = 0'925$$

# Probabilidad condicionada e independencia de sucesos

## Ejemplo 12

En un centro de secundaria el 50 % de los alumnos aprueba el Bachillerato. Se estima que si se presentasen todos los alumnos a las pruebas de Selectivo sólo suspenderían el 40 % y que un 30 % de los alumnos que aprobarían el Selectivo suspenden el Bachillerato. Con estos datos calcular la probabilidad de que un alumno que apruebe el Bachillerato apruebe el Selectivo.

Sea  $C$  el suceso “aprueba el Bachillerato” y  $S$  el suceso “aprueba el Selectivo”, por tanto,  $P(C) = 0'50$ ,  $P(\bar{S}) = 0'40$   $P(\bar{C}/S) = 0'30$ .

La probabilidad pedida es

$$P(S/C) = \frac{P(S \cap C)}{P(C)} = \frac{P(C/S)P(S)}{P(C)} = \frac{0'70 \cdot 0'60}{0'50} = 0'84$$

Salvador Naya

Introducción

Experimentos  
y sucesos

Definición de  
probabilidad

Probabilidad  
condicionada  
e  
independencia  
de sucesos

Teorema de  
Bayes

Variables  
aleatorias

Noción de  
variable  
aleatoria

Variable  
aleatoria  
discreta

Variable

# Probabilidad condicionada e independencia de sucesos

Salvador Naya

Introducción

Experimentos y sucesos

Definición de probabilidad

Probabilidad condicionada e independencia de sucesos

Teorema de Bayes

Variables aleatorias

Noción de variable aleatoria

Variable aleatoria discreta

Variable

## Regla del producto

Sean  $A_1, A_2, \dots, A_n$  sucesos tales que  $P\left(\bigcap_{i=1}^{n-1} A_i\right) > 0$ .

Entonces:

$$\begin{aligned} P\left(\bigcap_{i=1}^n A_i\right) &= \\ &= P(A_1)P(A_2/A_1)P(A_3/A_1 \cap A_2) \cdots P\left(A_n / \left(\bigcap_{i=1}^{n-1} A_i\right)\right) \end{aligned}$$

# Probabilidad condicionada e independencia de sucesos

Salvador Naya

Introducción

Experimentos y sucesos

Definición de probabilidad

Probabilidad condicionada e independencia de sucesos

Teorema de Bayes

Variables aleatorias

Noción de variable aleatoria

Variable aleatoria discreta

Variable

## Ejemplo 13

En relación con el ejemplo 10, si se eligen cuatro motores al azar, sin reemplazamiento, calcular la probabilidad de que el primer y el tercer motores elegidos sean defectuosos y los otros dos no.

$$\begin{aligned}P(D_1 N_2 D_3 N_4) &= \\&= P(D_1)P(N_2/D_1)P(D_3/D_1 N_2)P(N_4/D_1 N_2 D_3) = \\&= \frac{3}{10} \frac{7}{9} \frac{2}{8} \frac{6}{7} = \frac{1}{20} = 0'05\end{aligned}$$

# Probabilidad condicionada e independencia de sucesos

Salvador Naya

Introducción

Experimentos  
y sucesos

Definición de  
probabilidad

Probabilidad  
condicionada  
e  
independencia  
de sucesos

Teorema de  
Bayes

Variables  
aleatorias

Noción de  
variable  
aleatoria

Variable  
aleatoria  
discreta

Variable

## Definición

Dos sucesos  $A$  y  $B$  se dicen **independientes** si

$$P(A \cap B) = P(A)P(B)$$

o, equivalentemente,  $P(A/B) = P(A)$ , si  $P(B) > 0$ , o bien  $P(B/A) = P(B)$ , si  $P(A) > 0$ .

# Probabilidad condicionada e independencia de sucesos

Salvador Naya

Introducción

Experimentos y sucesos

Definición de probabilidad

Probabilidad condicionada e independencia de sucesos

Teorema de Bayes

Variables aleatorias

Noción de variable aleatoria

Variable aleatoria discreta

Variable

## Independencia de sucesos. Comentarios

- 1 La independencia de sucesos puede suponerse en algunas situaciones y deducirse del contexto del problema pero, en general, debe comprobarse experimentalmente.
- 2 No debe confundirse sucesos independientes con sucesos incompatibles.
- 3 Si  $A$  y  $B$  son sucesos independientes también lo son  $A$  y  $\bar{B}$ ,  $\bar{A}$  y  $B$  y  $\bar{A}$  y  $\bar{B}$ .
- 4 Los sucesos  $A_1, A_2, \dots, A_n$  son **mutuamente independientes** si  $P\left(\bigcap_{h=1}^k A_{j(h)}\right) = \prod_{h=1}^k P(A_{j(h)})$  para cualesquiera índices  $1 \leq j(1) < j(2) < \dots < j(k) \leq n$ .

# Probabilidad condicionada e independencia de sucesos

Salvador Naya

Introducción

Experimentos y sucesos

Definición de probabilidad

Probabilidad condicionada e independencia de sucesos

Teorema de Bayes

Variables aleatorias

Noción de variable aleatoria

Variable aleatoria discreta

Variable

## Ejemplo 14

Consideremos un sistema electrónico que consta de diez componentes que funcionan independientemente teniendo cada uno una probabilidad de fallo de 0'05. Calcular la *fiabilidad del sistema* (probabilidad de que el sistema funcione correctamente).

Si denominamos  $C_i$  al suceso “la componente  $i$ -ésima funciona correctamente”, donde  $i = 1, \dots, 10$ , con  $P(C_i) = 0'95$ , la fiabilidad del sistema es

$$P(C_1 C_2 \dots C_{10}) = 0'95^{10} = 0'598$$

# Probabilidad condicionada e independencia de sucesos

Salvador Naya

Introducción

Experimentos y sucesos

Definición de probabilidad

Probabilidad condicionada e independencia de sucesos

Teorema de Bayes

Variables aleatorias

Noción de variable aleatoria

Variable aleatoria discreta

Variable

## Ejemplo 14

Para aumentar la fiabilidad del sistema, se conectan en paralelo dos sistemas iguales al descrito. Calcular la fiabilidad del nuevo sistema.

Sea  $S_j$  el suceso “el sistema  $j$  funciona correctamente”, con  $j = 1, 2$ . Dado que  $P(S_j) = 0'598$ , la fiabilidad del nuevo sistema es

$$P(S_1 \cup S_2) = 0'598 + 0'598 - 0'598^2 = 0'838$$

Si conectásemos en paralelo tres sistemas como el primero, ¿cuál sería la fiabilidad del sistema resultante?

La fiabilidad de este último sistema es  $P(S_1 \cup S_2 \cup S_3) = 0'935$

# Teorema de Bayes

Salvador Naya

Introducción

Experimentos  
y sucesos

Definición de  
probabilidad

Probabilidad  
condicionada  
e  
independencia  
de sucesos

Teorema de  
Bayes

Variables  
aleatorias

Noción de  
variable  
aleatoria

Variable  
aleatoria  
discreta

Variable

## Teorema de las probabilidades totales

Sea  $A_1, A_2, \dots, A_n$  un sistema completo de sucesos, con  $P(A_i) > 0$  ( $i = 1, \dots, n$ ), y sea  $B$  un suceso cualquiera. Entonces

$$P(B) = \sum_{i=1}^n P(B/A_i)P(A_i)$$

# Teorema de Bayes

Salvador Naya

Introducción

Experimentos  
y sucesos

Definición de  
probabilidad

Probabilidad  
condicionada  
e  
independencia  
de sucesos

**Teorema de  
Bayes**

Variables  
aleatorias

Noción de  
variable  
aleatoria

Variable  
aleatoria  
discreta

Variable

## Ejemplo 15

En una escuela técnica el 50 % de los alumnos es de primer curso, el 30 % es de segundo y el 20 % de tercero. De la encuesta de evaluación de profesorado se sabe que el 60 % de los alumnos de primero tiene buena opinión del profesorado, al igual que el 70 % de los de segundo y el 75 % de los de tercero. Elegido un alumno al azar ¿cuál es la probabilidad de que tenga una buena opinión del profesorado?

# Teorema de Bayes

Salvador Naya

Introducción

Experimentos  
y sucesos

Definición de  
probabilidad

Probabilidad  
condicionada  
e  
independencia  
de sucesos

Teorema de  
Bayes

Variables  
aleatorias

Noción de  
variable  
aleatoria

Variable  
aleatoria  
discreta

## Ejemplo 15

Si consideramos el suceso  $B =$  “tener buena opinión del profesorado” y el sistema completo de sucesos formado por  $I =$  “ser de primero”,  $S =$  “ser de segundo” y  $T =$  “ser de tercero”, la probabilidad pedida es:

$$\begin{aligned}P(B) &= P(B \cap \Omega) = P(B \cap (I \cup S \cup T)) \\&= P(B \cap I) + P(B \cap S) + P(B \cap T) \\&= P(B/I)P(I) + P(B/S)P(S) + P(B/T)P(T) \\&= 0'6 \cdot 0'5 + 0'7 \cdot 0'3 + 0'75 \cdot 0'2 = 0'66\end{aligned}$$

# Teorema de Bayes

Salvador Naya

Introducción

Experimentos  
y sucesos

Definición de  
probabilidad

Probabilidad  
condicionada  
e  
independencia  
de sucesos

**Teorema de  
Bayes**

Variables  
aleatorias

Noción de  
variable  
aleatoria

Variable  
aleatoria  
discreta

Variable

## Ejemplo 16

En una estación de ITV (Inspección Técnica de Vehículos) hay dos equipos de inspección, el equipo A rechaza el 30 % de los coches inspeccionados y el equipo B no rechaza ningún coche. Si llegan tres coches a la estación y cada uno elige al azar uno de los dos equipos de inspección, ¿cuál es la probabilidad de que los tres coches superen la inspección?

# Teorema de Bayes

Salvador Naya

Introducción

Experimentos  
y sucesos

Definición de  
probabilidad

Probabilidad  
condicionada  
e  
independencia  
de sucesos

Teorema de  
Bayes

Variables  
aleatorias

Noción de  
variable  
aleatoria

Variable  
aleatoria  
discreta

Variable

## Ejemplo 16

Sean los sucesos  $A =$  “elegir equipo A”,  $B =$  “elegir equipo B” y  $S =$  “superar la inspección”, por el teorema de las probabilidades totales se obtiene

$$P(S) = P(S/A)P(A) + P(S/B)P(B) = 0'7 \cdot 0'5 + 1 \cdot 0'5 = 0'85$$

Denominemos  $S_i$  al suceso “el coche  $i$  supera la inspección”, con  $i = 1, 2, 3$ . Por la independencia de estos sucesos, la probabilidad pedida es

$$P(S_1 \cap S_2 \cap S_3) = 0'85^3 = 0'6141$$

# Teorema de Bayes

Salvador Naya

Introducción

Experimentos  
y sucesos

Definición de  
probabilidad

Probabilidad  
condicionada  
e  
independencia  
de sucesos

Teorema de  
Bayes

Variables  
aleatorias

Noción de  
variable  
aleatoria

Variable  
aleatoria  
discreta

Variable

## Teorema de Bayes

Sea  $A_1, A_2, \dots, A_n$  un sistema completo de sucesos, con  $P(A_i) > 0$  para  $i = 1, \dots, n$ , (**probabilidades a priori**) y sea  $B$  un suceso cualquiera, con  $P(B) > 0$ . Entonces, para  $j = 1, 2, \dots, n$ ,

$$P(A_j/B) = \frac{P(A_j B)}{P(B)} = \frac{P(B/A_j)P(A_j)}{\sum_{i=1}^n P(B/A_i)P(A_i)},$$

llamadas **probabilidades a posteriori**.

# Teorema de Bayes

Salvador Naya

Introducción

Experimentos  
y sucesos

Definición de  
probabilidad

Probabilidad  
condicionada  
e  
independencia  
de sucesos

**Teorema de  
Bayes**

Variables  
aleatorias

Noción de  
variable  
aleatoria

Variable  
aleatoria  
discreta

Variable

## Ejemplo 17

Se dispone de dos métodos para transmitir un mensaje, el método A transmite correctamente el 70 % de los mensajes y el método B el 90 %. Un día se elige un método al azar y se transmiten ocho mensajes comprobándose posteriormente que los dos primeros se han transmitido de forma incorrecta. ¿Cuál es la probabilidad de que se haya utilizado el método A? ¿Cuál es la probabilidad de que se haya utilizado el método B?

# Teorema de Bayes

## Ejemplo 17

Sean los sucesos  $A =$  “se utiliza el método A” y  $B =$  “se utiliza el método B”, con probabilidades  $P(A) = P(B) = 0'5$ . Denominemos  $M$  al suceso “se envían ocho mensajes, los dos primeros de forma incorrecta”, entonces

$$P(M/A) = 0'3^2 \cdot 0'7^6 = 0'01059$$

$$P(M/B) = 0'1^2 \cdot 0'9^6 = 0'00531$$

$$P(A/M) = \frac{0'5 \cdot 0'01059}{0'5 \cdot 0'01059 + 0'5 \cdot 0'00531} = 0'666$$

La probabilidad  $P(B/M)$  también puede calcularse utilizando Bayes, o directamente:  $P(B/M) = 1 - P(A/M) = 0'334$

Salvador Naya

Introducción

Experimentos  
y sucesos

Definición de  
probabilidad

Probabilidad  
condicionada  
e  
independencia  
de sucesos

Teorema de  
Bayes

Variables  
aleatorias

Noción de  
variable  
aleatoria

Variable  
aleatoria  
discreta

Variable

# Teorema de Bayes

Salvador Naya

Introducción

Experimentos  
y sucesos

Definición de  
probabilidad

Probabilidad  
condicionada  
e  
independencia  
de sucesos

**Teorema de  
Bayes**

Variables  
aleatorias

Noción de  
variable  
aleatoria

Variable  
aleatoria  
discreta

Variable

## Ejemplo 18

En un examen tipo test con cinco posibles respuestas, la probabilidad de que Juan sepa la respuesta es  $0'6$ , la probabilidad de que responda al azar es  $0'2$  y la probabilidad de que no responda es  $0'2$ . Si el estudiante respondió correctamente ¿cuál es la probabilidad de que realmente sepa la respuesta?

# Teorema de Bayes

Salvador Naya

Introducción

Experimentos  
y sucesos

Definición de  
probabilidad

Probabilidad  
condicionada  
e  
independencia  
de sucesos

Teorema de  
Bayes

Variables  
aleatorias

Noción de  
variable  
aleatoria

Variable  
aleatoria  
discreta

Variable

## Ejemplo 18

Sean los sucesos  $S$  = “Juan sabe la respuesta”,  $A$  = “Juan responde al azar” y  $N$  = “Juan no responde”, con probabilidades:  $P(S) = 0'6$ ,  $P(A) = 0'2$  y  $P(N) = 0'2$ . Sea  $C$  el suceso “Juan respondió correctamente”, se verifica que  $P(C/S) = 1$ ,  $P(C/A) = 1/5 = 0'2$  y  $P(C/N) = 0$ . Por el teorema de Bayes se obtiene:

$$P(S/C) = \frac{0'6 \cdot 1}{0'6 \cdot 1 + 0'2 \cdot 0'2 + 0'2 \cdot 0} = \frac{0'6}{0'64} = 0'9375$$

análogamente,  $P(A/C) = \frac{0'04}{0'64} = 0'0625$  y,  
 $P(N/C) = \frac{0}{0'64} = 0$

# VARIABLES ALEATORIAS

Salvador Naya

Introducción

Experimentos  
y sucesos

Definición de  
probabilidad

Probabilidad  
condicionada  
e  
independencia  
de sucesos

Teorema de  
Bayes

Variables  
aleatorias

Noción de  
variable  
aleatoria

Variable  
aleatoria  
discreta

Variable

## Noción de variable aleatoria

Sean un experimento aleatorio y su espacio muestral  $E$ .

- Una **variable aleatoria**  $X$  es una aplicación que asocia a cada elemento del espacio muestral un número real:

$$X : E \longrightarrow \mathbb{R}$$

- Una v.a. no es más que una forma de asociar un número real a cada uno de los posibles resultados de un experimento aleatorio.

# VARIABLES ALEATORIAS

Salvador Naya

Introducción

Experimentos  
y sucesos

Definición de  
probabilidad

Probabilidad  
condicionada  
e  
independencia  
de sucesos

Teorema de  
Bayes

VARIABLES  
ALEATORIAS

Noción de  
variable  
aleatoria

Variable  
aleatoria  
discreta

Variable

## Ejemplos

**Urna:** Se elige al azar una bola de una urna que contiene 2 bolas blancas y 1 bola negra. Una v.a. asociada a este experimento podría ser:

$$\begin{aligned} X : E = \{B, N\} &\longrightarrow \mathbb{R} \\ B &\longrightarrow 0 \\ N &\longrightarrow 1 \end{aligned}$$

**Dos dados:** Se lanzan dos dados. La función que asocia a cada resultado el máximo de los puntos obtenidos es una v.a.:

$$\begin{array}{cccccccc} E = & \{(1, 1), & (1, 2), & \dots, & (1, 6), & (2, 1), & \dots, & (6, 6)\} \\ \downarrow & \downarrow & \downarrow & \dots & \downarrow & \downarrow & \dots & \downarrow \\ \mathbb{R} & 1 & 2 & \dots & 6 & 2 & \dots & 6 \end{array}$$

## Noción de variable aleatoria

- La **función de distribución**  $F$  de una v.a.  $X$  es la función que asigna a cada n° real  $x$  la probabilidad de que la v.a. tome un valor menor o igual que  $x$ ; esto es, la probabilidad acumulada hasta dicho n° real:

$$\begin{aligned} F : \mathbb{R} &\longrightarrow \mathbb{R} \\ x &\longrightarrow F(x) = P(X \leq x) \end{aligned}$$

- La función de distribución nos proporciona toda la estructura probabilística asociada a la v.a. Por tanto, cualquier pregunta relacionada con dicha estructura puede ser contestada a partir de la función de distribución.

# Variables aleatorias

## Variable aleatoria discreta

Una **v.a. discreta** es aquella que toma un número finito de posibles valores, o bien un número infinito pero numerable (esto es, tantos valores como elementos tiene el conjunto de los números naturales).

## Ejemplos

- Las v.a. presentadas en los dos ejemplos anteriores (Ürnaz "Dos dados") son discretas.
- La v.a. que asocia a cada día el  $n^{\circ}$  de monitores producidos en una fábrica es discreta.
- Lanzamos una moneda sucesivamente hasta que obtengamos çara". La v.a. que asocia a cada resultado del experimento el  $n^{\circ}$  de lanzamientos realizados es discreta.

Salvador Naya

Introducción

Experimentos  
y sucesos

Definición de  
probabilidad

Probabilidad  
condicionada  
e  
independencia  
de sucesos

Teorema de  
Bayes

Variables  
aleatorias

Noción de  
variable  
aleatoria

Variable  
aleatoria  
discreta

Variable

## Variable aleatoria discreta

Sea  $X$  una v.a. discreta que toma los valores  $x_1 < x_2 < \dots$ .

- La **función de masa de probabilidad**  $p_X$  de dicha v.a. es la función que asigna a cada valor  $x_i$  la probabilidad de que la v.a. lo tome; esto es:

$$\begin{aligned} p_X : \{x_j\} &\longrightarrow \mathbb{R} \\ x_i &\longrightarrow p_i = P(X = x_i) \end{aligned}$$

- Se verifica que:
  - $p_i \geq 0, i = 1, 2, \dots$
  - $\sum_i p_i = 1$

# VARIABLES ALEATORIAS

## Ejemplo: Urna

$$\begin{aligned} p_X : \{0, 1\} &\longrightarrow \mathbb{R} \\ 0 &\longrightarrow 2/3 = P(\{B\}) \\ 1 &\longrightarrow 1/3 = P(\{N\}) \end{aligned}$$

$$F(x) = \begin{cases} P(\phi) = 0, & \text{si } x < 0 \\ P(\{B\}) = 2/3, & \text{si } 0 \leq x < 1 \\ P(E) = 1, & \text{si } x \geq 1 \end{cases}$$

**Observación:** La función de distribución de una v.a. discreta puede escribirse como:

$$F(x) = \sum_{i / x_i \leq x} p_i$$

Salvador Naya

Introducción

Experimentos  
y sucesos

Definición de  
probabilidad

Probabilidad  
condicionada  
e  
independencia  
de sucesos

Teorema de  
Bayes

Variables  
aleatorias

Noción de  
variable  
aleatoria

Variable  
aleatoria  
discreta

Variable

# VARIABLES ALEATORIAS

Salvador Naya

Introducción

Experimentos  
y sucesos

Definición de  
probabilidad

Probabilidad  
condicionada  
e  
independencia  
de sucesos

Teorema de  
Bayes

Variables  
aleatorias

Noción de  
variable  
aleatoria

Variable  
aleatoria  
discreta

## Variable aleatoria continua

Una **v.a. continua** es aquella que toma valores en un conjunto continuo (toda la recta real, un intervalo, o la unión de varios intervalos).

## Ejemplos

- La v.a. que asocia a cada persona su peso es continua.
- La v.a. que asocia a cada ejecución de un programa informático el tiempo utilizado es continua.

# VARIABLES ALEATORIAS

Salvador Naya

Introducción

Experimentos  
y sucesos

Definición de  
probabilidad

Probabilidad  
condicionada  
e  
independencia  
de sucesos

Teorema de  
Bayes

VARIABLES  
ALEATORIAS

Noción de  
variable  
aleatoria

Variable  
aleatoria  
discreta

## Variable aleatoria continua

La **función de densidad**  $f$  asociada a una v.a. continua  $X$  es el límite de los histogramas (de área 1) construidos con observaciones de dicha v.a. El límite se toma sobre el número de observaciones (que debe tender a  $\infty$ ) y sobre la amplitud de cada clase (que debe tender a 0). La gráfica de  $f$  describe el reparto de densidad de probabilidad de la v.a. continua  $X$ .

Se verifica que:

- $f(x) \geq 0, \forall x \in \mathbb{R}$ .
- $\int_{-\infty}^{\infty} f(x) dx = 1$ .
- $F(x) = P(X \leq x) = \int_{-\infty}^x f(t) dt$ .
- $P(a \leq X \leq b) = \int_a^b f(x) dx = F(b) - F(a)$ .
- $f(x) = F'(x)$  en los puntos  $x$  en los que  $F$  es derivable.  
En los que no lo es, se toma  $f(x) = 0$ , por convenio.

# VARIABLES ALEATORIAS

## Ejemplo

Si la función de distribución de una v.a.  $X$  es

$$F(x) = \begin{cases} 0, & \text{si } x < 0 \\ \frac{1}{2}x, & \text{si } 0 \leq x \leq 2 \\ 1, & \text{si } x > 2 \end{cases}$$

entonces su función de densidad es

$$f(x) = \begin{cases} \frac{1}{2}, & \text{si } 0 < x < 2 \\ 0, & \text{si } x \notin (0, 2) \end{cases}$$

Además,  $P\left(\frac{1}{4} \leq X \leq 1\right) = F(1) - F\left(\frac{1}{4}\right) = \frac{1}{2} \cdot 1 - \frac{1}{2} \cdot \frac{1}{4} = \frac{3}{8}$ .

Equivalentemente,

$$P\left(\frac{1}{4} \leq X \leq 1\right) = \int_{\frac{1}{4}}^1 f(x) dx = \int_{\frac{1}{4}}^1 \frac{1}{2} dx = \frac{3}{8}.$$

Salvador Naya

Introducción

Experimentos  
y sucesos

Definición de  
probabilidad

Probabilidad  
condicionada  
e  
independencia  
de sucesos

Teorema de  
Bayes

Variables  
aleatorias

Noción de  
variable  
aleatoria

Variable  
aleatoria  
discreta

Variable

## Ejemplo

Si la función de densidad de una v.a.  $X$  es

$$f(x) = \begin{cases} x, & \text{si } 0 < x \leq 1 \\ 2 - x, & \text{si } 1 < x < 2 \\ 0, & \text{si } x \notin (0, 2) \end{cases}$$

entonces su función de distribución es

$$F(x) = \int_{-\infty}^x f(t) dt = \begin{cases} 0, & \text{si } x \leq 0 \\ \frac{x^2}{2}, & \text{si } 0 < x \leq 1 \\ 2x - \frac{x^2}{2} - 1, & \text{si } 1 < x < 2 \\ 1, & \text{si } x \geq 2 \end{cases}$$

# VARIABLES ALEATORIAS

Salvador Naya

Introducción

Experimentos y sucesos

Definición de probabilidad

Probabilidad condicionada e independencia de sucesos

Teorema de Bayes

VARIABLES ALEATORIAS

Noción de variable aleatoria

Variable aleatoria discreta

Variable

## Esperanza y varianza de una variable aleatoria

La **esperanza matemática** o **media** de una v.a.  $X$  se define como:

- $E(X) = \mu = \sum_i x_i p_i$ , si  $X$  es discreta.
- $E(X) = \mu = \int_{-\infty}^{\infty} x f(x) dx$ , si  $X$  es continua.

La **varianza** de una v.a.  $X$  se define como:

- $Var(X) = \sigma^2 = \sum_i (x_i - \mu)^2 p_i$ , si  $X$  es discreta.
- $Var(X) = \sigma^2 = \int_{-\infty}^{\infty} (x - \mu)^2 f(x) dx$ , si  $X$  es continua.

Se tiene que  $Var(X) = E\left((X - \mu)^2\right) = E(X^2) - \mu^2$ , esto es:

- $Var(X) = \sum_i x_i^2 p_i - \mu^2$ , si  $X$  es discreta.
- $Var(X) = \int_{-\infty}^{\infty} x^2 f(x) dx - \mu^2$ , si  $X$  es continua.

La **desviación típica** es la raíz cuadrada positiva de la varianza.

# VARIABLES ALEATORIAS

Salvador Naya

Introducción

Experimentos y sucesos

Definición de probabilidad

Probabilidad condicionada e independencia de sucesos

Teorema de Bayes

VARIABLES ALEATORIAS

Noción de variable aleatoria

Variable aleatoria discreta

Variable

## Esperanza y varianza de una variable aleatoria

La esperanza y la varianza verifican:

- $E(X_1 + X_2 + \dots + X_n) = E(X_1) + E(X_2) + \dots + E(X_n)$ .
- $E(aX + b) = aE(X) + b$ , donde  $a, b \in \mathbb{R}$ .
- $Var(X_1 + X_2 + \dots + X_n) = Var(X_1) + Var(X_2) + \dots + Var(X_n)$ , si las v.a.  $X_i$  son independientes\*.
- $Var(aX + b) = a^2 Var(X)$ , donde  $a, b \in \mathbb{R}$ .

\*Las v.a.  $X$  e  $Y$  se dice que son **independientes** si lo son cualquier par de sucesos definidos por ambas. Ejemplo de tales sucesos podría ser  $\{X \leq 5\}$  e  $\{Y \neq 3\}$ . Esta definición se puede generalizar a más de dos variables.

## Distribución binomial

Supongamos que un experimento es realizado  $n$  veces, de modo que las  $n$  pruebas o realizaciones son independientes.

Denominemos "ÉXITO.<sup>a</sup>" un suceso de interés asociado al experimento, y sea  $p$  su probabilidad de ocurrencia.

La v.a.  $X$ ="Número de ÉXITOS en las  $n$  pruebas realizadas" se dice que tiene una **distribución binomial** de parámetros  $n$  y  $p$  ( $X \in B(n, p)$ ).

- Su función de masa de probabilidad viene dada por

$$p_X(k) = P(X = k) = \binom{n}{k} p^k (1-p)^{n-k}, \quad k = 0, 1, \dots, n.$$

- Su media es  $\mu = np$  y su varianza es  $\sigma^2 = np(1-p)$ .

# VARIABLES ALEATORIAS

Salvador Naya

Introducción

Experimentos  
y sucesos

Definición de  
probabilidad

Probabilidad  
condicionada  
e  
independencia  
de sucesos

Teorema de  
Bayes

Variables  
aleatorias

Noción de  
variable  
aleatoria

Variable  
aleatoria  
discreta

## Ejemplos

- Lanzamos 15 veces un dado, y consideramos la v.a.  $X$  = "Número de cuatros obtenidos". Esta v.a. tiene distribución  $B(15, 1/6)$  (el suceso ÉXITO es ".ºbtener cuatro"). El número medio de cuatros que se obtendrán es  $\mu = 15 \frac{1}{6} = 2,5$
- De una urna con 30 bolas blancas y 20 bolas negras extraemos 10 bolas con reemplazamiento, y consideramos la v.a.  $X$  = "Número de bolas negras extraídas". Esta v.a. tiene distribución  $B(10, 2/5)$  (el suceso ÉXITO es ".ºextraer bola negra"). El número medio de bolas negras que se extraerán es  $\mu = 10 \frac{2}{5} = 4$

## Don Anselmo y los pimientos de Padrón

Cierto restaurante dispone de un **gran lote** de pimientos de Padrón, de los cuales el 10 % son picantes. Don Anselmo toma en ese establecimiento 20 pimientos.

Denotando por  $X$  a la v.a. "número de pimientos que le pican a Don Anselmo", se tiene que su distribución es **aproximadamente** una  $B(20, 0,1)$ . Por tanto:

- La probabilidad de que no le pique ninguno es:

$$P(X = 0) \approx \binom{20}{0} 0,1^0 (1 - 0,1)^{20-0} = 0,9^{20} = 0,1216$$

## Don Anselmo y los pimientos de Padrón (continuación)

- La probabilidad de que le piquen dos o más es:

$$\begin{aligned}P(X \geq 2) &= 1 - P(X < 2) \\&= 1 - P(X = 0) - P(X = 1) \\&\approx 1 - 0,1216 - \binom{20}{1} 0,1^1 (1 - 0,1)^{20-1} \\&= 1 - 0,1216 - 20 \cdot 0,1 \cdot 0,9^{19} \\&= 1 - 0,1216 - 0,2702 = 0,6082\end{aligned}$$

- El número medio de pimientos que le picarán es  
 $\mu \approx 20 \cdot 0,1 = 2$

## Distribución normal

Se dice que una v.a.  $X$  tiene una **distribución normal** de parámetros  $\mu$  y  $\sigma$  ( $X \in N(\mu, \sigma)$ ) si su función de densidad es

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}, \forall x \in \mathbb{R}.$$

- La media de una v.a.  $X \in N(\mu, \sigma)$  es  $\mu$  y su varianza  $\sigma^2$ .
- La distribución  **$N(0, 1)$**  se denomina **normal estándar**, y la v.a. correspondiente se denota, en general, por  **$Z$** .
- Se verifica que

$$X \in N(\mu, \sigma) \iff Z = \frac{X - \mu}{\sigma} \in N(0, 1)$$

# VARIABLES ALEATORIAS

Salvador Naya

Introducción

Experimentos  
y sucesos

Definición de  
probabilidad

Probabilidad  
condicionada  
e  
independencia  
de sucesos

Teorema de  
Bayes

VARIABLES  
ALEATORIAS

Noción de  
variable  
aleatoria

Variable  
aleatoria  
discreta

Variable

## Don Anselmo y los percebes de Cedeira

Cada verano, a finales de Julio y/o comienzos de Agosto, Cedeira celebra su afamada "Festa do Percebe", en la que se reparten (previo pago) miles de raciones del mencionado crustáceo. En general, los organizadores indican que el contenido de cada ración pesa 250 gr, aunque en realidad, si denotamos por  $X$  a la v.a. "peso (en gr) de una ración", se tiene que  $X \in N(250, 25)$ . Don Anselmo va a comprar una ración.

La probabilidad de que pese menos de 225 gr es:

$$\begin{aligned} P(X < 225) &= P\left(\frac{X - 250}{25} < \frac{225 - 250}{25}\right) \\ &= P(Z < -1) = P(Z > 1) = 0,1587 \end{aligned}$$

# VARIABLES ALEATORIAS

## Don Anselmo y los percebes de Cedeira (continuación)

La probabilidad de que pese al menos 300 gr es:

$$\begin{aligned}P(X \geq 300) &= P\left(\frac{X - 250}{25} \geq \frac{300 - 250}{25}\right) \\ &= P(Z \geq 2) = 0,0228\end{aligned}$$

La probabilidad de que su peso esté entre 225 y 300 gr es:

$$\begin{aligned}P(225 \leq X \leq 300) &= P\left(\frac{225 - 250}{25} \leq \frac{X - 250}{25} \leq \frac{300 - 250}{25}\right) \\ &= P(-1 \leq Z \leq 2) \\ &= 1 - P(Z > 1) - P(Z > 2) \\ &= 1 - 0,1587 - 0,0228 = 0,8185\end{aligned}$$

Salvador Naya

Introducción

Experimentos  
y sucesos

Definición de  
probabilidad

Probabilidad  
condicionada  
e  
independencia  
de sucesos

Teorema de  
Bayes

Variables  
aleatorias

Noción de  
variable  
aleatoria

Variable  
aleatoria  
discreta

Variable

## Distribución normal

- Si  $X_1, X_2, \dots, X_n$  son v.a. independientes y con distribución  $N(\mu_1, \sigma_1), N(\mu_2, \sigma_2), \dots, N(\mu_n, \sigma_n)$ , respectivamente, y  $a_1, a_2, \dots, a_n$  son números reales, entonces:

$$a_1 X_1 + a_2 X_2 + \dots + a_n X_n \in N\left(\sum_{i=1}^n \mu_i, \sqrt{\sum_{i=1}^n a_i^2 \sigma_i^2}\right)$$

- **Teorema Central del Límite:** Si  $X_1, X_2, \dots, X_n$  son v.a. independientes e idénticamente distribuidas (i.i.d.) con media  $\mu$  y desviación típica  $\sigma$ , entonces, para  $n$  "grande" ( $n > 30$ ), se tiene que

$$X_1 + X_2 + \dots + X_n \approx N(n\mu, \sigma\sqrt{n})$$

## Distribución normal

- A medida que  $n$  aumenta, se tiene que la distribución binomial  $B(n, p)$  se acerca más y más a una distribución normal. En la práctica, dicha aproximación se da por válida cuando  $n > 30$  y  $0,1 < p < 0,9$  (o  $np(1 - p) > 5$ ). Así, en estas condiciones, se tiene que  $X \in B(n, p)$  puede aproximarse mediante  $Y \in N(np, \sqrt{np(1 - p)})$ .
- Puesto que la binomial es una distribución **discreta** y la normal es **continua**, es conveniente aplicar la siguiente **corrección por continuidad**. De este modo, la aproximación de las probabilidades es más fina.

$$P(X = k) \approx P(k - 0,5 \leq Y \leq k + 0,5), \quad k = 0, 1, \dots, n$$

# VARIABLES ALEATORIAS

## Ejemplos

Nos disponemos a lanzar 180 veces un dado, y deseamos calcular la probabilidad de obtener entre 25 y 40 cuatros.

La v.a.  $X$  = "Número de cuatros obtenidos" tiene distribución  $B(180, 1/6)$ , que puede aproximarse ( $n > 30$  y  $0,1 < p < 0,9$ ) por la  $Y \in N(180 \cdot \frac{1}{6}, \sqrt{80 \cdot \frac{1}{6} \cdot \frac{5}{6}}) = N(30, 5)$ . Por tanto:

$$\begin{aligned} P(25 < X < 40) &= P(X = 26) + P(X = 27) + \dots + P(X = 39) \\ &= P(26 \leq X \leq 39) \\ &\approx P(26 - 0,5 \leq Y \leq 39 + 0,5) = P(25,5 \leq Y \leq 39,5) \\ &= P\left(\frac{25,5 - 30}{5} \leq \frac{Y - 30}{5} \leq \frac{39,5 - 30}{5}\right) \\ &= P(-0,9 \leq Z \leq 1,9) = 1 - 0,0287 - 0,1841 = 0,7872 \end{aligned}$$

Salvador Naya

Introducción

Experimentos y sucesos

Definición de probabilidad

Probabilidad condicionada e independencia de sucesos

Teorema de Bayes

Variables aleatorias

Noción de variable aleatoria

Variable aleatoria discreta

Variable